

Will patient feedback improve Quality and Outcome in Psychotherapy?

**Implementation and validation of the online
feedback system OQ®-Analyst in Norway:**

A multi-site RCT study

Ingunn Agathe Amble, MD

MODUMBAD 
- en kilde til liv



UiO  **Det medisinske fakultet**

Universitetet i Oslo

Research Institute, Modum Bad
Institute of Clinical Medicine, Faculty of Medicine,
University of Oslo

Oslo 2016

© **Ingunn Agathe Amble, 2016**

*Series of dissertations submitted to the
Faculty of Medicine, University of Oslo*

ISBN 978-82-8333-247-6

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard
Printed in Norway: 07 Media AS – www.07.no

CONTENTS	PAGE
PREFACE	7
SUMMARY	9
LIST OF PAPERS	10
ACKNOWLEDGEMENTS	11
LIST OF ABBREVIATIONS AND SOME DEFINITIONS	14
1.0 INTRODUCTION	16
<i>1.1 Background</i>	16
<i>1.2 Feedback systems</i>	16
1.2.1 The OQ®-Analyst System	17
<i>1.3 Feedback research</i>	20
1.3.1 Controlled outcome studies	20
1.3.2 Is Feedback effective for all patients?	21
1.3.3 The role of the warning signal	22
<i>1.4 The purpose of the study</i>	24
2.0 AIMS	25
3.0 METHODS	25
<i>3.1 Instruments and Feedback System</i>	25
3.1.1 The Outcome Questionnaire 45.2 (OQ®-45)	26
3.1.2 The Symptom Checklist 90-Revised (SCL-90-R)	27
3.1.3 The Inventory of Interpersonal Problems (IIP-64-C)	27
3.1.4 The Social Adjustment Rating Scale (SAS)	28
3.1.5 The OQ-45 Feedback System (OQ®-Analyst)	28

3.2 Procedures and Participants	29
3.2.1 Ethics	29
3.2.2 Translation	29
3.2.3 Study 1	30
3.2.3.1 <i>Non-clinical samples</i>	30
3.2.3.2 <i>Clinical samples</i>	31
3.2.4 Study 2 and 3	33
3.2.4.1 <i>Clinics</i>	33
3.2.4.2 <i>Data collection</i>	34
3.2.4.3 <i>Patients</i>	35
3.2.4.4 <i>Therapists</i>	37
3.3 Analyses	37
3.3.1 Study 1	37
3.3.1.1 <i>Reliability and validity</i>	37
3.3.1.2 <i>Factor analyses</i>	38
3.3.2 Study 2	39
3.3.2.1 <i>Effect of using OQ®-Analyst</i>	39
3.3.2.2 <i>Site effect</i>	39
3.3.2.3 <i>Effect of being a not-on-track case</i>	39
3.3.3 Study 3	40
3.3.3.1 <i>Session at which Feedback had an effect</i>	40
3.3.3.2 <i>Effect of a warning signal</i>	40

3.3.3.3 <i>The severity of the patients` distress level</i>	41
3.3.3.4 <i>Therapist survey</i>	41
4.0 SYNOPSIS OF PAPERS	41
4.1 <i>Paper 1</i>	41
4.2 <i>Paper 2</i>	44
4.3 <i>Paper 3</i>	46
5.0 DISCUSSION	49
5.1 <i>Main results</i>	49
5.2 <i>The main issues from Paper 1.</i>	49
<i>The psychometric properties of OQ®-45</i>	
5.2.1 Reliability and validity	49
5.2.2 The representativeness of the samples	50
5.2.3 National experienced well-being and non-clinical means of the OQ-Total Score	52
5.2.4 The factor structure	52
5.3 <i>The main issues from Paper 2.</i>	53
<i>The effect of using the OQ®-Analyst in Norwegian psychiatric clinics</i>	
5.3.1 Difference between sites	53
5.3.2 Does feedback influence the number of sessions?	54
5.3.3 On-track versus not-on track cases	54

5.4 The main issues from Paper 3.	55
<i>How and When Feedback works: Is it the Signal?</i>	
5.4.1 The high proportion of not-on-track cases	55
5.4.2 Effect of the warning signal	56
5.4.3 Initial distress level	57
5.4.4 Early effect of feedback	57
5.4.5 Effect of feedback for not-on-track vs on-track cases	58
5.4.6 Therapist survey	58
5.4.7 The OQ®-Analyst system	59
5.5 Strengths and Limitations	60
6.0 CONCLUSIONS	61
7.0 IMPLICATIONS	62
REFERENCES	64
APPENDIX	
<i>The Norwegian version of the Outcome Questionnaire-45.2</i>	75
<i>The Therapist Survey Questionnaire, Norwegian</i>	77
PAPER 1	
PAPER 2	
PAPER 3	

Preface

The project presented in this thesis has its origin in a wish and a need for designing a clinically relevant way to evaluate the outcome of psychotherapy in ordinary psychiatric clinics and settings in Norway. These thoughts were lively discussed at the Outpatient Clinic at Modum Bad Psychiatric Center, an ordinary public outpatient clinic where adult patients with a broad spectrum of psychiatric conditions are seeking treatment. We often asked ourselves the essential question: Does the therapy we provide to patients really help?

As therapists, it was crucial for us that if we should implement a tool or assessment procedure to monitor change in therapy, to be shared with patients and therapists, it had to be easy to use in a clinical meaningful, pedagogical, and time effective way. The information collected should be comprehensive enough, but as short as possible (necessary and sufficient), in addition to being reliable and valid.

One of our therapists had used a patient feedback instrument, the paper/pencil version the Outcome Questionnaire -45 (OQ®-45), in treatment processes for some years. This questionnaire, comprising 45 items regarding the patient`s experience of distress, including symptom distress, interpersonal functioning, and social role functioning, is an essential part of the Outcome Questionnaire System (OQ®-Analyst).

The patient fills in the questionnaire prior to each session, and the OQ®-Analyst delivers a report containing the OQ- scores for that session as well as a graph of the progress in therapy up to that point in time. In addition, it yields self- report of three risk areas (suicidality, drug abuse and acting out), various warnings about whether expected progress had been achieved or not, and a computer generated verbal report. This information is provided to the therapist and can be shared with the patient as well. The OQ®-Analyst is one of many patient feedback systems which involve routine outcome monitoring and the provision of feedback to therapists and/or patients with the goal of measuring outcome to improving the quality of

mental health care (Howard, Moras, Brill, Martinovich, & Lutz, 1996; Miller, Duncan, Sorrell, & Brown, 2005; Duncan, 2012; Kraus, Seligman, & Jordan, 2005; Pinosof et al., 2009; Barkham, Gilbert, Connell, Marshall, & Twigg, 2005; Barkham et al., 2001; Evans et al., 2002, 2000; Lutz et al., 2005; Coombs, Burgess, & Pirkis, 2006; Wing et al., 1998).

It became clear to us that the OQ®-Analyst system was suitable for our outpatient setting for evaluating outcome as well as for improving the quality of psychotherapy and decreasing the number of failing cases. As our goal was to assess the quality of care and improve it, we decided to implement the online version of this patient feedback system.

To implement this system in Norway, we needed to translate it into Norwegian and investigate the psychometric properties of the Norwegian version. Second, in order to investigate the effect of using this patient feedback system in ongoing psychotherapies in Norwegian inpatient and outpatient psychiatric clinics treating patients with moderate to severe dysfunction, we designed a randomized clinical trial to test the effects of feedback to therapists versus treatment as usual (TAU i.e., no feedback). To obtain large enough samples with a great variety of conditions and settings, we conducted a multicenter study including six psychiatric clinics. Third, if OQ®-Analyst should prove to be effective, we were curious to find out more about how it worked.

Summary

This thesis deals with the aspects of monitoring psychotherapy in ordinary psychiatric settings in Norway with the goal of improving the quality of mental health services. The research resulted in three published papers. The first paper described the psychometric properties of the Norwegian version of the OQ®-45.2 with data from clinical and nonclinical samples. The OQ®-45.2 was found to be psychometrically reliable and valid for the purpose of monitoring and evaluating patient progress in therapy. The second paper described the results of the clinical trial comparing feedback to therapist and patients about patient progress, using the OQ®-Analyst, to TAU (no feedback). It was found that feedback improved the outcome and the quality of mental services. The third paper investigated how the system worked to improve outcomes by examining various aspects of the OQ®-Analyst. It could not be conclusively demonstrated that receipt of a warning signal for cases in the feedback condition resulted in better outcomes than in cases unaware of being off track. It was found that the effect of using OQ®-Analyst occurred already in the third session, and that the graphs displaying patient progress and the discussion of progress with the patient were experienced as the most important aspects of the system.

Finding the system to be reliable, valid, and effective both in improving outcome and optimizing the psychotherapeutic treatment, it can be recommended for implementation on a large scale in the Norwegian psychiatric health care facilities.

List of papers

I Ingunn Amble, Tore Gude, Sven Stubdal, Tuva Øktedalen, Anne Marie Skjørten, Bror Just Andersen, Ole André Solbakken, Hanne H. Brorson, Espen Arnevik, Michael J. Lambert & Bruce E. Wampold

Psychometric Properties of the Outcome Questionnaire-45.2: The Norwegian Version in an international Context.

Psychotherapy Research. 2014: Volume 24, Issue 4, p 504-513

II. Ingunn Amble, Tore Gude, Sven Stubdal, Bror Just Andersen, & Bruce E. Wampold
The effect of implementing the Outcome Questionnaire-45.2 feedback system in Norway: A multisite randomized clinical trial in a naturalistic setting. Psychotherapy Research. 2015: Volume 25, Issue 6, p 669-677

III. Ingunn Amble, Tore Gude, Sven Stubdal, Pål Ulvenes, & Bruce E. Wampold

How and When Feedback Works in Psychotherapy: Is it the Signal?

Psychotherapy Research. Published online July 14th, 2015.

Acknowledgements

Many have contributed to making this thesis possible. I wish to thank them all.

Primarily I want to thank all the patients and therapists at the Modum Bad outpatient clinic, and at the Drammen and Bærum District Psychiatric Centers who have participated in this study, filling out the questionnaires and using the feedback reports in the therapy sessions.

Normdata were delivered from inhabitants of Modum community, students, members of Friskis and Sveltis Fitness Center in Hønefoss, and employees at health institutions in Norway. Thank you all.

I also want to thank, Extrastiftelsen (The Council of Mental Health), the Norwegian Health Department and Modum Bad for funding this study.

Most important, I would like to thank my two principal supervisors professor Tore Gude and professor and director of the Research Institute, Modum Bad, Bruce E. Wampold for their endless patience, encouragement, and support, and for pushing and pulling me along this long and winding road. They have guided me step by step, and given constructive and prompt feedback. Together with the director of Modum Bad, Ole Johan Sandvand, I sometimes think of them as my three musketeers who have stood with me through thick and thin during these years.

I am also grateful to my two assistant supervisors professor Theresa Wilberg and professor Egil W. Martinsen from the Faculty of Medicine, University of Oslo for their knowledge and support.

The former director of the Research Institute, Modum Bad, professor Leigh McCullough, initiated this project in 2008. It is sad she is not longer here to see the result of her enthusiasm and support. Professor Michael Lambert, Brigham Young University, Utah, one of the developers of the OQ®-45 Analyst system, has generously been available for discussions, training and inspiration.

Without my colleague and superior project coordinator, Sven Stubdal, this project would not have been possible. He has been responsible for data collection, registration, randomization, and logistic. He has been supervising, training the therapists and staff, encouraging, checking, discussing, and communicating with the IT staff in Norway and US. His efforts were invaluable.

I want to thank the co-authors, project coordinators, and clerical staff at the different participating sites for their positive attitude and logistic skills which made the project feasible: Bror Just Andersen, Anne Marie Skjørten, Hanne Brorson, Henrich Bille, Ole André Solbakken, Espen Arnevik, Siw Kristiansen, Erna Sørum, Erna N. Bråthen, Ana Maria Silva-Harper, and Heidi Ellefsen.

A special thanks to my colleagues Pål Ulvenes and Tuva Øktedalen for stimulating conversations, and their willingness and patience to share their time, and knowledge of statistical subtleties of mixed models and factor analyses.

There have been many data technical challenges associated to this project. A big applause to Bjørn Henning Jahren for all the problems he has solved. I am also grateful to Terje Grendal for always being available, friendly and calm.

The project emanated from the staff at the Modum Bad outpatient clinic, and they have been my main source of support. Thank you so much to my very hard working dear colleagues for their humor, staying power, consideration, and willingness to share experiences, victories and frustrations.

I want to thank assistant director Terje Tilden, and all my colleagues at the Research Institute at Modum Bad, for valuable advice and encouragement.

Finally, I want to thank the most important persons in my life; my husband, Tor Magnus, and our four children, Sigmund, Øivind, Gyda, and Taran (who are not children

anymore), for their persistent support and ability to make me realize that life has many aspects.

The most important inspiration all these years has been the patients and the belief and hope that the results from this study could contribute to better outcome of psychotherapy. I dedicate this thesis to them and all future patients and therapists who perhaps will experience that using a feedback system increases the quality of care.

List of abbreviations and some definitions

OQ®-45	The Outcome Questionnaire 45.2. A 45-item patient self-report instrument designed to assess experience of psychological distress, including symptom distress, interpersonal functioning, and contentment with social role functioning.
OQ®-Analyst	The OQ-45 Feedback System. Provides the therapist and patient with a report containing the OQ- scores for that session as well as a graph of the progress in therapy up to that point in time. In addition, it yields self-report of three risk areas (suicidality, drug abuse and acting out), various warnings about whether expected progress has been achieved or not, and a computer generated verbal report.
OQ-Total Score	The sum of all 45 items gives a Total Distress score ranging from zero to 180, with higher scores being indicative of greater levels of psychological distress.
CST	Clinical Support Tools. Provide clinicians with actionable feedback when clients are not making expected feedback (NOT). It helps to identify important factors which can influence the outcome of psychotherapy, ie therapeutic alliance, motivation for change, social support, and critical life events. CST was not used in the present study.
TAU	Treatment as usual
FB	Feedback
NFB	No-feedback
NOT	Not-on-track, the patient is not making expected progress.
OT	On-track, the patient is making expected progress.

RCI	Reliable Change Index. If the change score exceeds the Reliable Change Index (which is calculated based on the reliability of the measure), then the patient can be said to be "significantly improved", meaning that the observed change would be expected by chance alone at a probability of less than 5%. This means that the change is unlikely to be due to simple measurement unreliability.
Clinical cut-off	The Clinical Cutoff Score refers to a score that is presumed to represent the boundary between "normal" and the "clinical range" on an outcome measure.
SCL-90-R	The Symptom Checklist 90-Revised
IIP-64-C	The Inventory of Interpersonal Problems
SAS	The Social Adjustment Rating Scale
COMPASS	the COMPASS integra Outpatient tracking System
PCOMS	the Partners for Change Outcome Management System
TOP	the Treatment Outcome Package
STIC	the Systemic Therapy Inventory of Change
Core-OM	the Clinical Outcomes in Routine Evaluation – Outcome Measure
HoNOS	the Health of the Nation Outcome Scales

1.0 Introduction

1.1 Background

There is an established knowledge within the field that psychotherapy in general is effective (Lambert, 2013; Lambert & Ogles, 2004; Fonagy, Roth, & Higgitt, 2005).

Individuals with psychiatric problems or symptoms benefit from therapy compared to those who do not receive therapy. A meta-analysis of research outcome studies has shown that about 75% of patients improve (Hansen, Lambert, & Forman, 2002; Wampold & Imel, 2015).

However, in routine outpatient psychotherapy practice less than 50% of the cases improve and about 5-10% actually get worse, being treatment failures or non-responders (Hansen & Lambert, 2003; Hansen, Lambert & Foreman, 2002; Lambert & Ogles, 2004; Mohr, 1995).

There is definitely a need to improve the quality of services. This knowledge urged us to ask: Could the use of OQ®-Analyst increase the number of patients benefitting from psychotherapy, and perhaps decrease the number of patients deteriorating?

The current efforts to improve the quality of services involve the dissemination of evidence-based treatments into practice settings (e.g., Shafran et al., 2009). Another method to improve the quality of services, involves what has been called practice-based evidence (Barkham, Hardy, & Mellor-Clark, 2010; Lambert, 2010; Miller, Duncan, & Hubble, 2005). Practice-based evidence involves utilizing evidence from practice to improve the quality of service. One way in which practice-based evidence is used, is to provide feedback from patients about their progress to therapists. The feedback can be shared with the actual patient.

1.2 Feedback systems

Feedback from patients about their progress has been established as a viable means to improve the quality of psychotherapy services (Berking, Orth, & Lutz 2006; Harmon et al., 2006; Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004; Lambert et al., 2002; Whipple et al., 2003). The rationale for these systems has been that the patients' perspective on their

therapy is of crucial importance, and that this information is important for the therapist to know. There are several essential features of feedback models. First, patient's distress is measured over the course of therapy, typically every session. Second, the patient's progress is compared with normative trajectories of change, which typically are adjusted for initial levels of mental distress. Third, the information about patient progress in comparison to normative trajectories is available to the therapist (and the patient) to guide how the therapy is conducted.

Several feedback systems have been developed, including COMPASS (Howard, Moras, Brill, Martinovich, & Lutz, 1996), the Outcome Questionnaire Analyst system (OQ®-Analyst) (Lambert et al., 2004; Lambert, Kahler, Harmon, & Burlingame, 2011), the Partners for Change Outcome Management System (PCOMS) (Miller, Duncan, Sorrell, & Brown, 2005; Duncan, 2012), the Treatment Outcome Package (TOP) (Kraus, Seligman, & Jordan, 2005), and the Systemic Therapy Inventory of Change (STIC) (Pinsof et al., 2009) in the United States; the Clinical Outcomes in Routine Evaluation – Outcome Measure (Core-OM) (Barkham, Gilbert, Connell, Marshall, & Twigg, 2005; Barkham et al., 2001; Evans et al., 2002, 2000; Lutz et al., 2005) in the United Kingdom; and the Health of the Nation Outcome Scales (HoNOS) in the United Kingdom, Australia and New Zealand (Coombs, Burgess, & Pirkis, 2006; Wing et al., 1998).

1.2.1 The OQ®-Analyst System

The OQ®-Analyst uses the Outcome Questionnaire-45 (OQ®-45) to measure the patient's distress level for quantifying and evaluating patient progress. Lambert and Burlingame introduced the OQ®-45 in 1996 (Lambert, Burlingame et al., 1996). The OQ®-45 is a 45-item patient self-report instrument designed to assess the patient's experience of psychological distress (i.e., symptoms), interpersonal functioning, and contentment with social role functioning (Lambert et al., 1996). The OQ®-45 is typically given prior to the first

and each subsequent therapy session, in either a paper/pencil or an electronic format, and takes 5-10 minutes to complete. The sum of all items gives a Total Distress score (OQ-Total Score), ranging from zero to 180, with higher scores being indicative of greater levels of psychological distress.

The psychometric properties of the OQ®-45 have been studied extensively in the United States (Beretvas, Kearney, & Barón, 2003; Bludworth, Tracey, & Glidden-Tracey, 2010; Chapman, 2003; Kim, Beretvas, & Sherry, 2010; Lambert et al., 1996; Lambert et al., 2011; Mueller, Lambert, & Burlingame, 1998). Due to promising results in the US, the OQ®-45 has also been widely used around the world. Mental health systems as well as cultural values related to mental health vary from one country and context to another (Bhugra & Bhui, 2007), and , therefore, it should not be assumed that properties of the instrument are invariant across cultures. However, variations do not necessarily invalidate the instrument; indeed, such variations might as well inform us about differences in mental health services and cultural contexts.

To date, the OQ®-45 has been translated into more than 20 languages and psychometric analyses have been conducted in Germany (Lambert, Hannover, Nisslmüller, Richard, & Kordy, 2002), the Netherlands (de Jong et al., 2007), Italy (Chiappelli, Coco, Gullo, Bensi, & Prestano, 2008; Lo Coco et al., 2008), Sweden (Wennberg, Philips & de Jong, 2010), and China (Qin & Hu, 2008). Several conclusions can be drawn from these psychometric studies. First, the OQ®-45, in the US and in other countries, with translated versions, appears to be a reliable (stable and internally consistent) instrument. Second, the validity of the OQ®-45, as evidenced by relatively robust correlations with other instruments, appears to be strong across the various countries. However, the means of the OQ-Total Score of non-clinical and clinical samples varies among countries. As a consequence, the clinical

cut off and the Reliable Change Index (RCI) of the OQ®-45 differs as well, as discussed below.

Finally, confirmatory factor analyses reveal that the three factor solution hypothesized by Lambert has not been consistently confirmed, although the fit of the three factors to the data could be characterized as “moderately good.” Other solutions, including one to ten factors, as well as bi-level solutions (e.g, specific factors nested within one general factor) have also been investigated (Beretvas et al., 2003; Bludworth et al., 2010; Chapman, 2003; Chiappelli et al., 2008; de Jong et al., 2007; Kim et al., 2010; Mueller, 1998). Therefore, evidence points to that the most optimal use of OQ®-45 is to use OQ-Total Scores rather than the subscales.

The OQ®-Analyst software provides the therapist and patient with a feedback report showing the patient`s session by session progress. The report includes a graph showing the curve of the OQ-Total Scores from session to session, superimposed onto curves showing the expected progress of a patient with the same initial distress level as the actual patient, as well as a signal with a corresponding message. The signal is red or yellow if the patient`s progress is poorer than expected, and these cases are designated as signal cases or not-on-track cases. A green signal indicates that the patient`s progress is as expected, and a white signal indicates that the patient`s distress is below the clinical cut-off, both of which are defined as on-track cases. The report also includes the patients` response to five risk items regarding suicidal thoughts, drug/alcohol abuse, and acting out.

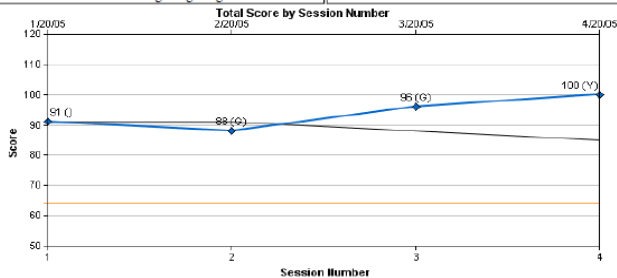
Figure1. An example of a Clinician Feedback Report generated by the OQ®-Analyst software for a patient who has completed the OQ®-45.2.

(From the OQ®-Analyst Documentation Files, <https://maine.oqanalyst.com/oqa/files/>)

Name: An. Adult, 2	ID: 24059	Alert Status: Yellow
Session Date: 4/20/2005	Session: 4	Most Recent Score: 100
Clinician: Clinician, Randy	Clinic: South Clinic	Initial Score: 91
Diagnosis: Depression		Change From Initial: No Reliable Change
Algorithm: Empirical		Current Distress Level: Moderately High

Most Recent Critical Item Status:	
8. Suicide - I have thoughts of ending my life. Frequently	
11. Substance Abuse - After heavy drinking, I need a drink the next morning to get going. Sometimes	
26. Substance Abuse - I feel annoyed by people who criticize my drinking. Frequently	
32. Substance Abuse - I have trouble at work/school because of drinking or drug use. Frequently	
44. Work Violence - I feel angry enough at work/school to do something I might regret. Rarely	

Subscales	Current	Output. Norm	Comm. Norm
Symptom Distress:	56	49	25
Interpersonal Relations:	27	20	10
Social Role:	17	14	10
Total:	100	83	45



Graph Label Legend:

(R) = Red: High chance of negative outcome (Y) = Yellow: Some chance of negative outcome
(G) = Green: Making expected progress (W) = White: Functioning in normal range

Feedback Message:

The rate of change the patient is making is less than expected. This patient may end up with no significant benefit from therapy. It is recommended that you be alert to the possible need to improve the therapeutic alliance, reconsider the client's readiness for change and the need to renegotiate the therapeutic contract, intervene to strengthen social supports, or possibly alter your treatment plan by intensifying treatment, or shift intervention strategies. Continue to carefully monitor treatment progress.

REMEMBER, THE USER IS SOLELY RESPONSIBLE FOR ANY AND ALL DECISIONS AFFECTING PATIENT CARE. THE OOB-A IS NOT A DIAGNOSTIC TOOL AND SHOULD NOT BE USED AS SUCH. IT IS NOT A SUBSTITUTE FOR A MEDICAL OR PROFESSIONAL EVALUATION. RELIANCE ON THE OOB-A IS AT USER'S SOLE RISK AND RESPONSIBILITY. SEE LICENSE FOR FULL STATEMENT OF RIGHTS, RESPONSIBILITIES & DISCLAIMERS.

This system seemed to meet most of our requirements, in addition, which we considered important, it contained items on suicidal risk and drug abuse. The online version is easy to use for both patient and therapist, and it produces immediate online reports and signals providing information about whether the patient was making expected progress (on-track) or not (not-on-track). Such information could be used by the therapist to decide whether the therapy should be adjusted to achieve better results.

1.3 Feedback research

1.3.1 Controlled outcome studies

Of the earlier mentioned patient Feedback systems, only the OQ®-Analyst and the PCOMS have been investigated with randomized clinical trials to determine their effect on outcome. Meta-analyses of trials that have compared feedback to treatment-as-usual (no

feedback) have shown that the OQ®-Analyst and PCOMS feedback systems improve the quality of care, primarily by improving outcomes for cases that are at risk for deterioration, which are often called “signal cases” or “not-on-track” (Lambert & Shimokawa, 2011a, 2011b; Shimokawa, Lambert, & Smart, 2010). The OQ®-Analyst is the most widely tested system and the results have been particularly promising (Lambert & Shimokawa, 2011b; Shimokawa et al., 2010). In a comprehensive review of six trials of the OQ-45, Shimokawa et al. (2010) concluded, “The current state of evidence appears to support the efficacy and effectiveness of feedback interventions in enhancing treatment outcomes” (p. 298). Although the evidence is convincing, there are some caveats that limit the veridicality and generalizability of the results. One of the issues is that the research on the OQ-45 trials reviewed by Shimokawa et al., were conducted in collaboration with the team who originally developed the system. Second, with one exception (viz., Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004), all patients in these trials were university college students receiving services at the same university college counseling center, a private Mormon (Latter Day Saints; LDS) university (Shimokawa et al., 2010).

However, in the years after the meta-analyses mentioned above, additional trials using the OQ®-Analyst have been published. Five trials, involving patients who were not university students, including outpatients as well as speciality care (including substance abuse outpatients, eating disorder patients, and psychosomatic inpatients) found that feedback using the OQ®-Analyst was effective (Crits-Cristoph et al., 2012; De Jong et al., 2014; Probst et al., 2013; Simon et al., 2013; Simon, Lambert, Harris, Busath, & Vazquez, 2012).

1.3.2 Is Feedback effective for all patients?

Despite the evidence showing that feedback from the OQ®-Analyst is beneficial, there is a need for independent replications of the results, particularly with more severely distressed patients. To our knowledge, only two OQ®-Analyst randomized clinical trials have been

conducted outside the United States. In a large study with more than 400 patients with diverse disorders from three outpatient clinics in the Netherlands, de Jong and colleagues (de Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012) found no significant effect of OQ®-Analyst feedback, even for patients who were not-on-track. However, not-on-track patients whose therapists actively used the feedback did achieve better outcomes than patients of the same therapists for whom feedback was not used and it appeared that feedback was more effective with relatively shorter treatment i.e, less than 35 sessions (de Jong et al., 2014). A study with 252 patients in psychosomatic in-patient treatment in Germany concluded that using OQ®-Analyst feedback improved the outcome for patients at risk for treatment failure (not-on-track) (Probst et al., 2013), but not for the on-track-patients.

It is theorized that feedback helps to resolve problems for patients who are predicted to be treatment failures. Some evidence suggests this is the case as the effect appears to occur most robustly for not-on-track cases ($d = .53$) (Shimokawa et al 2010). For patients making expected progress (on-track), feedback does not appear to improve outcomes (Lambert, Whipple, Smart, Vermeersch, & Nielsen, 2001; Lambert et al., 2002; Whipple et al., 2003; Harmon et al., 2005; Slade, Lambert, Harmon, Smart, & Bailey, 2008; de Jong et al., 2012; Probst et al., 2013; Simon, Lambert, Harris, Busath, & Vazquez, 2012; Crits-Christoph et al., 2012)

1.3.3 The role of the warning signal

Thus, it would appear that receiving a *warning signal* might play the most critical role in feedback. Indeed, Probst et al. (2013) found that trajectories of not-on-track patients were similar in Feedback (FB) and No Feedback (NFB) conditions until the therapist in the FB condition was signaled that the patient was off track. From that point forward the outcomes in patients in the FB condition improved significantly more than those in the NFB condition

(i.e., in the condition where the therapist did not receive any feedback, including the warning signal).

Probst et al.'s (2013) is consistent with the fact that in most of the studies feedback proved to be effective for the not-on-track patients only. If this were true, trajectories prior to the occurrence of the first warning signal would not be affected by receiving information about patient progress, as was found by Propst et al. (2013). Furthermore, if receiving the warning signal is the critical issue, then off-track patients in the FB condition would have better outcomes than in the NFB condition *after* the point in therapy where the patient went off track (i.e., received the first warning signal) (Propst et al., 2013). This explanation is bolstered by the fact that therapists, in the absence of any feedback, tend to be unaware of when their patients are off-track (Hannan et al., 2005). Thus it appears that therapists need to be made aware of when their patients go off track through the use of warning signals so that they can alter therapy in a more helpful direction for the patient.

An alternative explanation for how feedback works to improve outcomes involves the integration of the patient progress information into treatment. On the one hand, therapists receiving information about patient progress may pay little attention to the graph and reports until a warning signal is provided. On the other hand, therapists might discuss the feedback report with the patient as part of the therapy procedure, independent of the occurrence of a warning signal. Such a procedure might communicate to the patient that improvement is most important and that the therapist is eager to adjust therapy to ensure the best outcome possible. In this respect, a discussion about the process and outcome of treatment might have a beneficial effect. Some systems, such as the PCOMS, emphasize integrating the information into the process of therapy by discussing the patient's progress in every session (Duncan et al. 2012), while the OQ®-Analyst leaves it to the therapist to share the report information with the patient or not.

It is also important to consider the fact that Hawkins et al. (2004) and Simon et al. (2013) found the effect of feedback to be similar for on-track and not-on-track patients. These studies have in common that the patients, *both* in the on-track and not-on-track conditions, had a relatively higher initial level of distress compared to other feedback studies. As well, in the other studies, where an effect was found only for the not-on-track cases, these cases had a higher initial level of distress than the on-track cases (Lambert et al., 2001; Lambert et al., 2002; Whipple et al., 2003; Crits-Christoph et al., 2012). These findings support the notion that feedback could be more important for patients with more severe dysfunction.

1.4 The purpose of the study

As the findings vary across studies, countries, settings and patients groups, we wanted to investigate what would be the effect of feedback in a Norwegian naturalistic sample, which contains rather severely distressed patients. It appeared useful to apply a stepwise approach to investigate the effect and outcome of using the OQ®-Analyst in Norway. First, we had to (a) test the psychometrics of the Norwegian version of the OQ®-45, (b) investigate by means of a RCT whether receiving feedback would improve outcomes of psychotherapy in Norwegian naturalistic psychiatric settings, and (c) explore the how feedback works by identifying the effective aspects of the system.

We know that the implementation process can be very difficult in routine care, even when the effects of feedback are clear. The main problem is often that the therapists stop using the systems, even though they know that it is effective and can help their patients to obtain a better therapy outcome. Walter, Cleary, and Rey (1998) also showed that two-thirds of the therapists were not willing to use feedback, even if it would lead to demonstrably better outcomes.

As we wanted to know how the therapists experience using the OQ®-Analyst, we therefore, at the end of the study, conducted a survey of participating therapists asking if they had used the system as expected, and what aspects of the feedback system they rated as most important. After the study was closed, OQ®-Analyst was implemented as routine tool in Nov 2013 at the Psychiatric Outpatient Clinic at Modum Bad Psychiatric Center.

2.0 Aims

Based on the extant knowledge about feedback systems, we established three aims of this study.

1. To investigate the psychometric properties of the Norwegian version of OQ®-45 (Study 1)
2. To estimate the effect of using OQ®-Analyst Feedback System in Norwegian naturalistic psychotherapeutic settings compared to treatment as usual (i.e., no feedback) (Study 2).
3. To determine which elements of the OQ®-Analyst to be responsible for the benefits of using it, how and when it is effective, and what the experiences of therapists using OQ®-Analyst are (Study 3).

3.0 Methods

3.1 Instruments and Feedback System

The primary instrument used in this study is the Outcome Questionnaire 45.2. To validate the Norwegian translation, we also administered the Symptom Checklist 90-Revised (SCL-90-R), the Inventory of Interpersonal Problems (IIP-64-C), and the Social Adjustment Rating Scale (SAS) in the psychometric study. These instruments are described below.

3.1.1 The Outcome Questionnaire 45.2 (OQ®-45)

The OQ®-45 is an instrument developed for the purpose of following closely patient progress (Lambert et al., 1996). The OQ®-45 is a 45-item patient self-report instrument designed to assess experience of psychological distress, interpersonal functioning, and contentment with social role functioning, assessed with a five point Likert scale (0=never, 1=rarely, 2=sometimes, 3=frequently, 4= almost always). Nine of the items are reversed to limit the likelihood of response bias. The OQ®-45 is typically given prior to the first and each subsequent therapy session, in either paper/pencil or electronic formats, and takes 5-10 Minutes to complete. The sum of all items gives an OQ-Total Score, ranging from zero to 180, with higher scores being indicative of greater levels of psychological distress.

The clinical cutoff (the cut-off between being in the clinical and the normal range) is estimated based on the mean of the OQ-Total Scores from two representative samples, one clinical, and one non-clinical (Jacobson & Truax, 1991). The Reliable change index (RCI) represents the number of scale points needed to determine if a change in OQ-Total Score is due to real change or is due to the unreliability of the instrument (i.e, the standard error of measurement). The RCI is calculated using the standard error of measurement to estimate the range of chance variation (Jacobson & Truax, 1991), calculated from data from a psychiatric outpatient sample (see Paper 1 for statistical details).

To examine the concurrent validity of the OQ®-45, including the OQ-Total Score as well as the three subscales, we administered the Symptom Checklist 90-Revised (SCL-90R) the Inventory of Interpersonal Problems (IIP-64-C) and the Social Adjustment Rating Scale (SAS) We expected the SCL-90R to correlate with the Symptom Distress (SD) subscale, the IIP-64-C to correlate with the Interpersonal Relation(IR) scale, and the SAS to correlate with the Social Role Functioning (SR) scale.

3.1.2 The Symptom Checklist 90-Revised (SCL-90-R)

The Symptom Checklist 90-Revised (SCL-90-R) (Derogatis, 1977; Derogatis & Melisaratos, 1983) provides information about the client's experience of psychological symptoms as measured by nine symptom areas, and an index for overall symptom burden. The questionnaire contains 90 items rated on a Likert scale from 0=not at all to 4=very much, examining how the patient has experienced the symptoms in the past two weeks, including the present day. Only the index for overall symptom burden (GSI) of the Norwegian edition (Pedersen & Karterud, 2004) was used in this study. The internal consistency of the original version is reported to be .77 to .90 and the test-retest reliability .78 to .90 (Derogatis, 1994). The Norwegian version has been found to be internal consistent ($\alpha = .96$), and stable (one week test-retest $r = .83$) (Vassend, Lian, & Andersen, 1992).

3.1.3 The Inventory of Interpersonal Problems (IIP-64-C)

The IIP-64-C (Alden, Wiggins, & Pincus, 1990) is a shortened version of the The Inventory of Interpersonal Problems (Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, 1988). The IIP-64-C contains 64 statements describing common interpersonal problems. The patient is instructed to consider each item regarding whether it has been a problem with respect to any significant person in the patient's life and is assessed with Likert scale (0=not at all to 4=extremely). The IIP-64 contains eight subscales and a global score. The global score of the Norwegian version of IIP-64 was used in the present study. The original version has adequate psychometric properties with Cronbach's α of .72 to .85 and test-retest r of .78; (Horowitz, Alden, Wiggins, & Pincus, 2000). The Norwegian IIP has similar psychometric properties, with α ranging from .82 to .94 for the sub-scales and r of .98 for a test-retest correlation for the Total IIP score (Monsen, Hagtvet, Havik, & Eilertsen, 2006; Pedersen, 2002).

3.1.4 The Social Adjustment Rating Scale (SAS)

The Social Adjustment Rating Scale (Weissman & Bothwell, 1976) is a 54 item questionnaire that measures performance over the past two weeks in six areas of functioning: work; domestic or academic responsibilities; social and leisure activities; relationship with extended family; marital role as a spouse; parental role; and membership in the family unit. Each item is rated on a five-point scale (1 to 5), with the higher score being indicative of greater distress or impairment. The psychometric properties of the original version are Cronbach's α of .74 and test-retest r of .80 (Edwards, Yarvis, Mueller, Zingale, & Wagman, 1978)

3.1.5 The OQ-45 Feedback System (OQ@-Analyst)

The OQ-Analyst software (OQ@-Analyst) provides the therapist and patient with a report showing a graph of the patient's trajectory up to the current session. The report displays a comparison of the given patient's rate of improvement with expected rates of improvements based upon samples of patients with the same initial level of distress. The feedback report includes a colored signal determined by the empirical method (i.e., by comparing progress to normative trajectories; see Lambert et al., 2011). A white signal indicates that the patient's score is similar to people in a state of normal functioning (i.e., below the clinical cutoff of 62), and termination of therapy could be considered. A green signal indicates that the progress is as expected but there is still need for more treatment. A yellow signal is given when there is concern about the patient's progress or that a positive outcome is in doubt. A red signal indicates serious concern about the final outcome and that there is risk of deterioration unless changes are made. The signals occur based on calculations of tolerance intervals. The tolerance intervals were calculated for the expected mean OQ-45 Total Score at each session. If, at any session following intake, the OQ-45 Total Score for a patient is within the 68% tolerance interval, then therapy is

proceeding as anticipated for this particular patient, and a green message can be given as feedback for the therapist to proceed as usual. If the same OQ-Total Score falls above of the upper 68% tolerance interval (upper 16%), but is still within the upper bound of the 80% tolerance interval, the patient is beginning to deviate more than one standard deviation from what is expected of a typical person at this point in therapy. Then the therapist would receive a yellow message as a warning to attend to this patient's progress. This one standard deviation unit approximates a 14-point increase in the OQ-Total Score, which is the criterion for reliable change in the US version. Scores above the upper bound of the 80% tolerance interval, generate a red signal. If a patient has one or more red or yellow signals (warning signals) during therapy, the case is defined as a not-on-track or signal case.

3.2 Procedures and Participants

3.2.1 Ethics

The participating patients signed an informed, written consent. The study was approved by the Regional Ethical Research Committee in the South of Norway June 12th, 2009, and given an expanded approval to include patients from other clinics and a revised consent form June 30th 2011. The study was also approved by the Data Protection officer at Modum Bad representing The Norwegian Data Protection Authority.

3.2.2 Translation

In order to investigate the psychometric properties of a Norwegian version of the OQ[®]-45, we established a translation process involving a group of three native Norwegian therapists and researchers who separately translated the English OQ[®]-45, the feedback reports, including all the different feedback messages into Norwegian. When discrepancies occurred, the translations were discussed until consensus was obtained in the group. This consensual version was translated back to English by an experienced American researcher and therapist living and working in Norway, in cooperation with the group of Norwegian therapists and the

developer of the OQ[®]-45 Michael J. Lambert in order to obtain semantic equivalence (Flaherty, Gaviria, Pathak, & Mitchell, 1988). The final version was obtained by discussion in the extended group until consensus was reached.

3.2.3 Study 1.

The next step was to establish the psychometric properties of the Norwegian version of OQ[®]-45. We collected OQ-Total Scores from representative non-clinical and clinical samples (Table 1). To ensure heterogeneity of the non-clinical population, participants were sampled from four different sources: inhabitants in a rural Norwegian community, members of a fitness center in a small city, students from different academic fields at a college and a university, and employees in mental health institutions. The clinical samples were obtained from urban and rural psychiatric routine care clinics, two inpatient and four outpatient clinics.

3.2.3.1 Non-clinical samples. The Total Non-Clinical Sample (TNCS, N=338), which was used for calculating the non-clinical mean, Cronbach's α , and clinical cut-off, consisted of three subsamples:

Non-clinical sample 1 (NCS1). OQ-45 Total Scores were obtained from adult residents in a rural Norwegian community with a population of approximately 12000. The sampling was conducted by posting two OQ[®]-45 Questionnaires, a stamped response envelope, and an information letter emphasizing the inclusion criteria: Age above 18 years and not having a psychiatric diagnosis or being in current psychiatric treatment, in 300 random mailboxes. Information about the project was also given in the local newspaper and television channel. We received 99 envelopes with a total of 127 completed forms, which represents an estimated response rate of approximately 27 percent (assuming that national rate of 1.6 adults per household). In addition 59 members of a local fitness center completed the OQ[®]-45 in a city of approximately 30,000 inhabitants.

Non-clinical sample 2 (NCS2). OQ®-45 scores were collected from 46 students from different departments at the University of Oslo and a nearby college.

Non-clinical sample 3 (NCS3). A total of 106 employees at different psychiatric institutions throughout Norway completed the OQ®-45. Of these 57 persons filled it in twice, at a one week interval, providing data for calculating a test-retest correlation.

There were no significant differences in OQ-45 Total Scores among the (a) rural residents, (b) fitness center members, (c) students, and (d) health institutions employees ($F(3, 334) = 2.19, p = .089$), and consequently we merged subsamples to increase sample size.

3.2.3.2 Clinical samples. The Total Clinical Sample (TCS) (N=560) consisted of three subsamples:

Clinical sample 1 (CS1). Data from 280 patients were collected as initial OQ-Total Scores from three outpatient clinics in the southeast of Norway. These data were also used for calculating means, reliability, clinical cut-off, and the Reliable Change Index.

Clinical sample 2 (CS2). In this sample 184 patients (others than in CS1) from the same outpatient clinics as mentioned above filled in OQ®-45 only once during their course of treatment.

Clinical Sample 3 (CS3). In this sample 96 patients from three inpatient clinics completed the OQ®-45 at the beginning of therapy.

In the outpatient samples used in this study (CS1), about one-third of patients were diagnosed with anxiety disorders, about one-third with affective disorders, about 15 percent with unspecified diagnoses, about 6 percent with some type of behavioral disorder, and about 4 percent with eating disorders; less than 4 percent of the patients were diagnosed with a psychotic disorder, personality disorder, or substance use disorder. Of course, the validity of diagnoses in clinical settings must be considered with caution. Because we wanted to collect

a naturalistic patient sample, no diagnostic exclusion criteria were used although patients for whom their therapists assessed that the patient was unable to understand the meaning of OQ®-45 were not invited to participate.

Clinical sample 4 (CS4). This sample consisted of 61 patients (a subsample of CS1) who completed the Symptom Checklist 90-Revised (SCL-90-R) and The Inventory of Interpersonal Problems Circumplex (IIP-64-C) as well as the OQ®-45 prior to their intake session. The SCL-90-R and IIP-64-C were completed some days to weeks before the OQ®-45; time between the administrations was examined in the analyses.

Clinical sample 5 (CS5). This sample consisted of 54 patients (another subsample of CS 1) who completed the Social Adjustment Scale (SAS) as well as the OQ®-45 at the beginning of therapy. Participants who left more than eight of the OQ®-45 items unanswered were excluded. In total, only 14 participants had more than eight missing items. In case of missing values, a mean for the remaining scale items was calculated and replaced the missing values. However, missing values were not replaced in the analyses making use of data on item level (viz. factor analyses and analyses of internal consistency). Of the 132 participants who had one to eight missing values, 64 % had one missing value, 19 % had two missing values, 8 % had three missing values, and 8 % had more than three but less than eight missing values.

Table 1 Non Clinical and Clinical Samples demographics

Sample	Abbreviation	Analyses	N total used in analysis	% Female	Mean Age	Deleted due to > 8 missing items
Non Clinical Samples						
1. Community and Fitness Center	NCS1	Mean, clinical cut-off, Cronbach's α .	186	70.5	51.33	5
2. Students	NCS2	Mean, clinical cut-off, Cronbach's α .	46	65.1	23.60	2
3. Employees of Health Institutions	NCS3	Mean, clinical cut-off, Cronbach's α , Test-retest **	106	75.0	49.45	2
Total Non Clinical Sample	TNCS	Mean, clinical cut-off, Cronbach's α .	338	70.7	46.69	9
Clinical Samples						
1. Outpatients Initial scores	CS1	Mean, clinical cut-off, Cronbach's α . Reliable Change index. Factor analyses	280	67.9	35.8	0
2. Outpatients Not initial scores	CS2	Factor analyses, Cronbach's α .	184	74.3	33.3	3
3. Inpatients	CS3	Factor analyses, Cronbach's α .	96	56.1	39.4	2
Total Clinical Sample	TCS	Factor analyses, Cronbach's α .	560	68.2	34.5	5
Clinical samples for validation						
1. Outpatients	CS4	Concurrent validity with SCL-90 & IIP	61	70.5	36.9	0
2. Outpatients	CS5	Concurrent validity with SAS	54	70.4	37.0	0

*Not all of the individuals filled in Age and Gender

** 57 out of 106 filled in OQ[®]-45 twice with one week interval

3.2.4 Study 2 and 3

Data collected from participating patients and therapists, as described below, was used in Study 2 and 3.

3.2.4.1 Clinics. For the purpose of obtaining a sufficient number of participants, this project was designed as a multisite study. Twelve clinics were invited to participate, and six

consented. Two of the participating clinics were inpatient clinics and four were outpatient clinics, of which one was a substance abuse clinic. The clinics signed a cooperation agreement based on the research protocol and agreed to implement the randomization procedures and contribute data to the project.

At each site, one information meeting (1.5 hours) and one therapist training course (2.5 hours) were conducted. At one of the clinics, where the project leader was employed, the therapist participation was obligatory (Clinic 1, Table 2), while at the other clinics therapists' participation was voluntary.

3.2.4.2 Data collection. Because the aim of the trial was to examine the OQ®-Analyst in the context of naturalistic clinical settings, the therapists were not instructed to conduct treatment according to specific protocols or clinical treatment methods. A monthly OQ-forum was arranged to address logistical, practical, and clinical issues, and to evaluate and share experiences with the use of OQ®-Analyst. The project leader, the project coordinators, and two IT-consultants were available for personal supervision by telephone or by mail for all the participating therapists and office staffs during the entire project period.

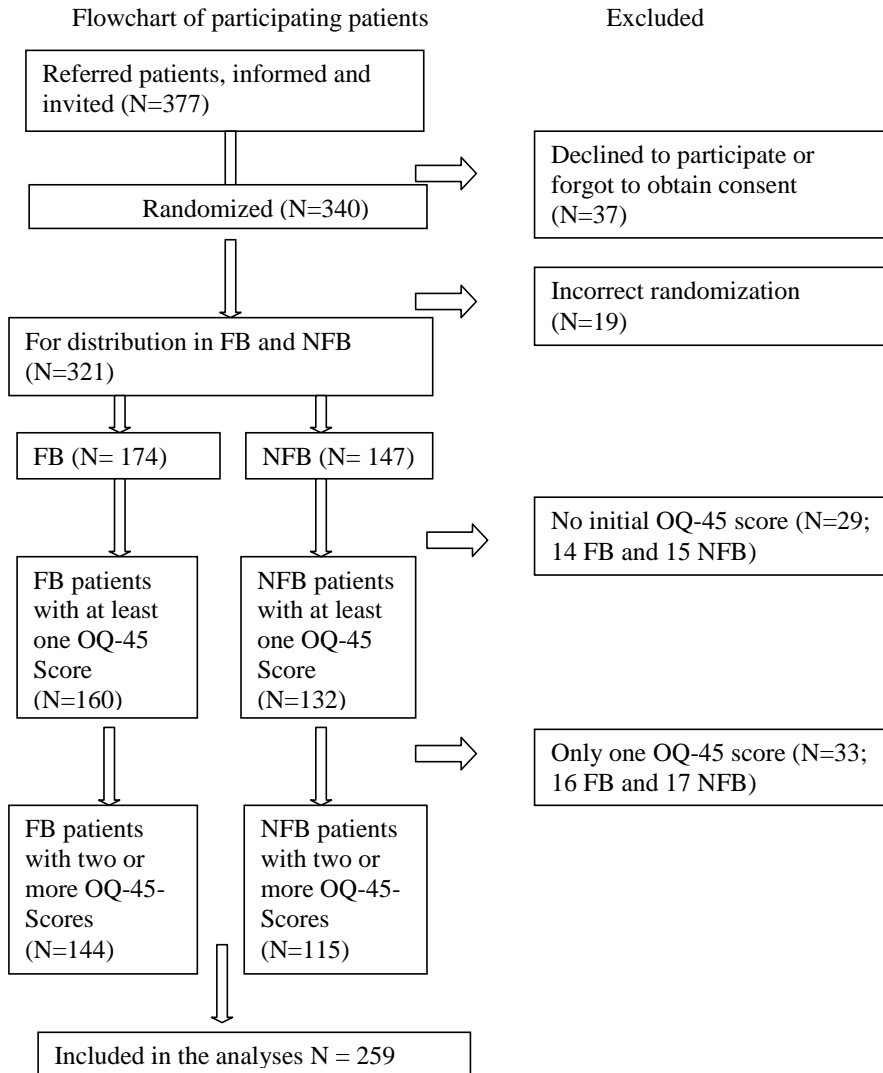
After having signed and returned an informed consent, the patients were randomized into the FB or NFB condition in blocks of eight and by gender. All the patients, in both conditions, filled in the OQ®-45 online prior to each session. They were informed about their randomization condition after having completed the initial score. In this study reports and alerts were generated by the OQ®-Analyst software on a session-by-session basis each time a patient completed the measure. For the patients in the FB condition, the therapists were instructed to consider the Feedback report, show it to the patient every session, and discuss the report when useful or necessary. Although we do not know the extent to which therapists showed the reports to the patients, we have classified the trial as comparing feedback to both therapists and patients to no feedback (i.e., opposed to trials that restrict feedback to the

therapist only). No further specific instructions on how to use the Feedback reports were given, but it was emphasized that the Feedback Reports should be opened, studied, and used in a clinical useful way. The feedback reports on the NFB patients could not be opened by anyone other than the project leader. The data collection period was June 2010 to September 2013.

3.2.4.3 Patients. During the inclusion period (June 2010 to September 2013), a total of 377 adult patients referred to outpatient or inpatient treatment, were invited to participate in the present study, as part of the clinic's intake procedures. No exclusion criteria other than the inability to complete the OQ[®]-45, as determined by the therapist, were established. Of the eligible patients, 37 patients did not return a signed consent, either because they declined to participate, the therapist determined they were unable to complete the OQ[®]-45, or the therapist forgot to ask them to complete the consent. The mean age of the patients was 35.8 years (*SD*: 11.66, Range 18-65) and 68% were female (see Table 2).

The 340 patients (280 of them were also contained in the CS1 in Study 1) initially consenting to participate were randomized to the FB or the NFB condition. At one of the sites, 19 patients were not randomized correctly and were therefore excluded from the sample, leaving 321 to be distributed to the two conditions. To be included in the analyses, a patient was required to have completed the outcome measure for a minimum of two sessions representing the first and any subsequent session. Over the course of the study, 14 patients in the FB and 15 in the NFB condition never started to score the OQ[®]-45; 16 patients in the FB and 17 in the NFB condition completed the OQ[®]-45 only once. Therefore these patients were excluded, leaving 259 patients to constitute the sample to be investigated, distributed in 144 in the FB and 115 in the NFB condition. Patient inclusion and exclusion is presented in Figure 1.

Figure 1



Patients seen in the clinics had a wide range of diagnoses and co-morbidity, except in the substance abuse clinic where all patients had a substance abuse disorder as their primary diagnosis. All patients were diagnosed by their therapist using International Classification of Diseases (ICD-10). No inter-rater reliability procedures were established. The primary diagnoses, from the most frequent, were as follows: various affective disorders (47%), anxiety disorders (33%), behavioral disorders (mainly Attention Deficit/Hyperactivity

Disorder, 7%), substance abuse (all from the substance abuse clinic; 4%), eating disorders (4%), personality disorders (3%), schizophrenia (1%), and no diagnosis (1%).

3.2.4.4 Therapists. Of the 135 licensed therapists employed at the six involved centers, 45 participated in the study. They provided a variety of theoretically guided treatments, including cognitive-behavioral, psychodynamic, and eclectic orientations. The mean age in the therapist group was 48.5 (*SD*: 9.54, Range 32-66), the mean years of experience as a therapist was 11.4 years (*SD*: 6.81, range: 1-31), 63 % of the therapists were female and represented various professions, including psychologists (41%), medical doctors (18% psychiatrists and 15% MD in training), psychiatric nurses (16%), and others (10%). The participating therapists had between one and 39 patients in treatment, distributed on both FB and NFB condition with a mean number of patients of 6.1 (*SD*: 7.17).

Table 2 Participating clinics, patients and therapists

Clinic	Clinic Description	Number Patients	Number FB	Number NFB	Female	Number and % of the employed therapists participating
1	Outpatient	127	67	60	73%	13/14 (93%)
2	Outpatient	40	26	14	63%	13/60 (22%)
3	Outpatient	51	30	21	77%	8/24 (33%)
4	Substance abuse, Outpatient	11	5	6	46%	5/16 (31%)
5	Inpatient	7	3	4	57%	2/12 (17%)
6	Inpatient	23	13	10	52%	4/9 (44%)
Total		259	144	115	69%	45/135 (33%)

Note. FB = Feedback Condition, NFB = No feedback condition

3.3 Analyses

3.3.1 Study 1

3.3.1.1 Reliability and validity. In order to establish the psychometric properties of the Norwegian version of the OQ[®]-45 we estimated the internal consistency (Cronbachs α) in the

clinical and the non-clinical samples (TCS, N = 560 and TNCS, N = 338, Table 1), and the test-retest correlation in the sample with at least two assessments (NCS1, N = 57) (Pearson's r).

To examine the concurrent validity of the OQ-45 Total Score as well as relevant subscales were correlated with the referent instrument using the clinical sample 3 (CS3, N=61) for the SCL-90R and IIP-64C, and clinical sample 4 (CS4, N=54) for the SAS. Divergent validity was assessed by correlating the OQ[®]-45 subscales with instruments not expected to cover the same area. Clinical cut-off and the reliable change index (RCI) were calculated according to the Jacobson and Truax method (Jacobson & Truax, 1991). In order to test the relationship between distress and well-being, Non-Clinical Means for the various countries were correlated with the National Experienced Well-being Index (New Economic Foundation, 2012). We correlated the means of the OQ-Total Scores from the non-clinical samples in different countries with the same country's National Experienced Well-being Index to examine whether OQ-45 Total Scores for the non-clinical samples in various countries reflect the well-being of their citizens.

3.3.1.2 Factor analyses. To test the consistency of the three-factor model of OQ[®]-45, we conducted a confirmatory factor analysis of Lambert's three factor structure using LISREL (8.8). According to Jöreskog and Sörbom (1993) a robust maximum likelihood estimation procedures suggested by Satorra and Bentler (1990) is preferred for data obtained from Likert type format and representative of an ordinal scale. The subprogram PRELIS 2.8 was used to estimate the respective covariance and their asymptotic covariance matrices assuming alternative parameterization for the factor analytic models (Jöreskog, 2005; Sörbom & Jöreskog, 1999). General guidelines suggest the following criteria for confirming a hypothesized factor structure: RMR < 0.10, RMSEA < 0.05, CFI and NNFI > 0.9, a non-

significant χ^2 , and $\chi^2/df < 2$ (Browne & Cudeck, 1992; Byrne, 2009; Hu & Bentler, 1998; MacCallum, Browne, & Sugawara, 1996).

For a number of reasons, we did not modify the hypothesized factor structure to better fit the data, test other models, or conduct exploratory factor analyses. First, the subscales are rarely used in research or clinical work, and if they are, Lambert's three subscales are used. Second, such strategies capitalize on sample characteristics and are contradictory to a confirmatory modeling strategy. Third, the factor structures have varied across countries and adding another "exploratory" structure would not contribute to our understanding of the OQ[®]-45 or augment the clinical application of the OQ[®]-45.

3.3.2 Study 2

3.3.2.1 Effect of using OQ[®]-Analyst. For the purpose of examining the effects of using OQ[®]-Analyst feedback in Norwegian psychiatric inpatient and outpatient clinics, the patients were randomly assigned to FB or NFB as described in Procedures. The primary analysis examined the effects of feedback (i.e., FB versus NFB) in a general linear model, where the final OQ-Total Score was the dependent variable and the first OQ-Total Score was entered as a covariate.

3.3.2.2 Site effect. To determine possible site effect, the site the patient was treated in was modeled as a random factor. To test whether effects of feedback differed among the clinics, the interaction of clinics and FB/NFB was tested.

3.3.2.3 Effect of being a not-on-track case. Because the effects of feedback have been found to be particularly apparent with not-on-track cases in other studies (i.e., cases that received a yellow or red flag during the course of therapy), we examined not-on-track cases and the interaction of not-on-track cases and feedback. To assess the size of the effect for feedback, we calculated Cohen's *d* (Cohen, 1988; Durlak, 2009) for the post-test OQ-Total Scores comparing FB to NFB for the entire sample.

3.3.3 Study 3

The findings from Study 2 inspired us to investigate the effect of the warning signals more thoroughly, and also look into at which point in therapy the effect of feedback appeared significant.

3.3.3.1 *Session at which Feedback had an effect.* To investigate the point at which feedback had an effect, we selected patients who had eight or more sessions and graphed their OQ-Total Score from session to session for both conditions. As well, we conducted independent sample t-tests for estimating change in the OQ-Total Scores from the one time point to the other.

3.3.3.2 *Effect of a warning signal.* To examine the effects of the warning signal, we examined progress of therapy before and after the signal was delivered in the FB cases and after the point where a signal would have been delivered in the NFB cases. Thus, the patients who received a signal in the FB cases ($n = 59$) or would have received a signal in the NFB condition ($n = 51$) were selected. Longitudinal piecewise multilevel models were used to estimate the slopes of the OQ-Total Scores before and after the first warning signal was given (Bauer & Curran, nd). In these analyses, unstructured covariance models were used because sessions occurred in varying time patterns (e.g., the time between sessions varied both between and within patients) and because commonly used structures (e.g., AR and ARIMA) produced poorer fitting models. To accomplish these analyses, the session at which it was determined that the patient first went off track was identified for each patient, and the longitudinal data were centered at this point. First, an unconditional model with only slopes and intercept was examined. The second model was conditioned on whether feedback was given or not. Longitudinal piecewise multilevel models were used to estimate the slopes of the OQ-Total Scores before and after the first signal was given (Fitzmaurice, Laird, & Ware, 2004).

3.3.3.3 *The severity of the patients` distress.* With the empirical support for the conjecture that feedback is more important for more severely distressed patients, we performed a Multiple Regression Analysis with last OQ-Total Score as dependent variable in order to test the hypothesis that initial severity was related to the effect of feedback. First score (OQ-Total Scores of the initial session), condition (FB versus NFB), as well as the interaction of the first score and condition were independent variables.

3.3.3.4 *Therapist survey.* Finally, we conducted a survey of the participating therapists with questions about various aspects of the feedback system (OQ@-Analyst) assessed with Likert scale (1=to a very small extent to 5=to a very large extent), and calculated means.

4.0 Synopsis of papers

4.1 Paper 1: Psychometric Properties of the Outcome Questionnaire-45.2: The Norwegian Version in an international Context.

Objective: Monitoring of ongoing psychotherapy is crucially important to improving the quality of mental health care by detecting therapies being off track, which requires that the instrument used is psychometrically sound. This study investigates the psychometric properties of the Norwegian version of the Outcome Questionnaire 45.2 (OQ@-45) and situates the results in an international context.

Method: The OQ@-45 is a 45 item form designed to measure the patient`s distress level (OQ-Total Score) distributed on three sub-scales, covering symptom distress (SD), interpersonal relations (IR), and social role functioning (SR). Data from one clinical sample (N = 560) and one non-clinical sample (N = 338) were used to investigating reliability, cut-offs, factor structure, and correlation with well-being data. All results were compared to similar analyses in various international contexts.

Results: The data showed adequate reliability and concurrent validity. Concerning reliability, the internal consistency yielded a Cronbach's $\alpha = .93$ both in the clinical sample (N=560) and in the non-clinical sample (N=338). The test-retest correlation with one week interval from 57 of the participants in the non-clinical sample 1 (NCS1) yielded a Pearson's r of .85. To examine the concurrent validity of the OQ®-45, the subscale scores were correlated with referent instruments. The correlations for the referent instruments and the OQ®-45 subscale scores ranged from .59 to .68, with SCL-90 Global Symptom Index (GSI) .68 for the SD score, .50 for IIP and IR and .67 for the SAS and SR. Divergent validity was assessed by correlating the OQ®-45 sub-scales with non-referent instruments (e.g., SCL-90, a symptom measure, with SR). These divergent validity coefficients were nearly as large as the convergent validity coefficients, and in some instances, larger (see Table 3).

Table 3. Reliability and validity values

Test retest Pearson's r	.85	
Internal Consistency α	.93	
Concurrent validity correlations with Total Scale and Corresponding Subscale with correlations for Divergent Validity	Total	Subscale
SCL-90 Global Symptom index	.68	.64 (SD) .61 (IR) .60 (SR)
IIP	.59	.50 (IR) .52 (SD) .55 (SR)
SAS	.67	.70 (SR) .58 (SD) .53 (IR)
Clinical Cut-off	62	
Reliable Change index	16	

Note: SCL-90R= The Symptom Checklist 90-Revised, IIP = The Inventory of Interpersonal Problems, SAS = The Social Adjustment Rating Scale, SD = Symptom Distress Scale, IR = Interpersonal Relations Scale, SR = Social Role Scale.

The Clinical cut-off level between case/non-caseness was estimated according to the Jacobson and Truax method (Jacobson & Truax, 1991), based on the following equation with

the OQ-Total Score being 37.3 in the non-clinical sample (N=338), and 90.8 in the clinical outpatient sample (N=280). ,

$$\begin{aligned} \text{Cutoff} &= \\ & \frac{(M_{\text{Clinical}} * sd_{\text{NonClinical}}) + (M_{\text{NonClinical}} * sd_{\text{Clinical}})}{(sd_{\text{Clinical}} + sd_{\text{NonClinical}})} = \\ & \frac{(90.8 * 18.5) + (37.3 * 21.1)}{(21.1 + 18.5)} \\ & = 62.4 \end{aligned}$$

When checking for gender differences, slightly higher means in the OQ-Total Score in females (93.7, SD = 21.2) compared with males (87.7, SD = 20.8) ($t = 2.16$, $p = 0.03$) in the clinical sample (N=280) were found. In the non-clinical sample (N=338), no gender differences occurred. The difference in the clinical sample did not change the clinical cutoff level when tested for each gender separately.

The reliable change index (RCI), calculated from the Outpatient sample using the formula provided by Jacobson and Truax (Jacobson & Truax, 1991) was as follows

$$RCI = 1.96SD\sqrt{2(1-r)} = (1.96 * 21.1)\sqrt{2(1-0.926)} = 15.91$$

where SD was the standard deviation and r the Cronbach's α .

Within the total sample of outpatients assessed at the beginning of treatment (N=280), 91.4 % had an OQ-45 Total Score greater than the Norwegian cutoff of 62, thus being classified as clinical cases. On the other hand 90.8% of the non-clinical sample (N=338) had an OQ-45 Total Score less of 62 or less, and were thus correctly classified as non-cases.

Each subscale is meant to measure a specific domain of functioning and therefore the subscales should be relatively uncorrelated. The correlations among the subscales ranged from .51 to .66 (SD/IR = .61, SD/SR = .66, IR/SR = .51).

Testing Lambert's three-factor model to the data with a confirmatory factor analysis yielded coefficients that did not fully meet standard criteria (RMR= .092; RMSEA = 0.082; CFI = .92; NNFI = .92; Satorra-Bentler Scaled $\chi^2_{942} = 4448.0863$, $p < 0.01$; $\chi^2/df = 4.72$),

although it could be said that the fit was to some degree acceptable in that some coefficients were in the acceptable range. The correlation among the latent factors of SD, IR, and SR ranged from .57 to .75, demonstrating that the three factors were not independent.

The correlation between the non-clinical means and national experienced well-being, was -0.942, indicating the lower the OQ-Total Score non-clinical mean for individual countries, the greater the experienced well-being, which establishes that differences for the non-clinical means among countries might well reflect true differences in well-being for the citizens of these countries.

Conclusions: The reliability and validity of the instruments showed that the Norwegian version of the OQ@-45 had adequate psychometric properties. The means, clinical cut-offs, and the reliable change index, vary across countries. However, the means of the OQ-Total Score for nonclinical samples correlate highly with external values of national well-being, indicating that the OQ@-45 is a valid instrument internationally. The factor analyses in the present study do not confirm the hypothesized factor structure of the OQ@-45, but are similar to the results internationally.

4.2 Paper 2: The effect of implementing the Outcome Questionnaire-45.2 feedback system in Norway: A multisite randomized clinical trial in a naturalistic setting.

Objective: It has been claimed that the monitoring of ongoing psychotherapy is of crucial importance for improving the quality of mental health care. This study investigated the effect of using the Norwegian version of the patient feedback system OQ@-Analyst using the Outcome Questionnaire-45.2.

Method: Patients from six psychiatric clinics in Southern Norway (N = 259) were randomized to FB or NFB. The FB group comprised 144 patients, while the NFB group comprised 115 patients. Patients in both groups had OQ-Total Scores from the first and the

last session. Demographics, means and standard deviations of the OQ-Total Scores for the two conditions (FB and NFB), site, and percentage of not-on-track cases were recorded. To illustrate the implications of the feedback effect, we classified all patients as recovered, improved, unchanged, or deteriorated. The Clinical Cut-off and the Reliable Change Index were used to calculate how the patients in the FB and the NFB groups were distributed in these four different Outcome Groups. Patients were defined as recovered if their final OQ-Total Score was in the nonclinical range (i.e., below 62) and the patients achieved reliable change (i.e., more than 16 points of change from initial to final OQ-Total Score). Patients were classified as improved if their final OQ-Total Score exhibited reliable change but remained in the clinical range. If the OQ-Total Score increased by more than 16 points (i.e., reliable deterioration), the patient was classified as deteriorated. Patients who did not meet any of these criteria were defined as unchanged.

Results: The main effect of feedback was statistical significant ($F = 3.80, p = .027$), corroborating the hypothesis that feedback would improve the quality of services. To assess the size of the effect from feedback, Cohen's d (Cohen, 1988; Durlak, 2009) was calculated for the difference in OQ-Total Scores (between first and last score) comparing FB with NFB in the entire sample and was found to be 0.32. No statistical differences between FB and NFB in the number of sessions ($t(257) = 0.41, p > .50$) nor in the proportion of not-on-track cases ($\chi^2 = 0.12, p > .50$) occurred. A main effect for clinic and for not-on-track cases was found as expected; however no interaction effects between sites and FB/NFB as well as between not-on-track/on-track case and FB/NFB occurred indicating that neither clinic nor being a not-on-track case had a significant effect on the degree to which feedback improved outcomes. The percentages in each category based on reliable change in the FB and NFB conditions showed that FB increased the probability of improvement and reduced the probability of treatment failure.

Conclusions and implications: The main effect of feedback was statistically significant ($p = .027$), corroborating the hypothesis that feedback would improve the quality of services, although the size of the effect was small to moderate ($d = 0.32$). The benefits of feedback have to be considered against the costs of implementation.

4.3 Paper 3: How and when feedback works in psychotherapy: Is it the Signal?

Objective: Studies show that monitoring of ongoing psychotherapy improves the quality of mental health care, and that feedback about patients' progress has been established as a viable means. The essential feature of feedback models is that patient progress is measured continuously through therapy. This study investigated the effect of receiving a warning signal when a patient is not achieving expected improvement (not-on-track), monitored with the Norwegian version of the patient feed-back system OQ®-Analyst.

Method: Patients from six psychiatric clinics in Southern Norway ($N = 259$) were randomized into the FB or NFB conditions. To examine the effect of the warning signals, we selected the 59 patients in the FB condition and the 51 in the NFB condition who produced one or more signals. To investigate the point at which feedback had an effect, we selected the 124 patients who had eight or more sessions and graphed their outcomes for each condition. Therapists were surveyed at the end of the study upon their way of ranking the different aspects of using OQ®-Analyst. They responded to items on a 5 point scale (1 = to a very small extent to 5 = to a very large extent).

Results: To examine the effects of receiving a warning signal, an unconditional model with only slopes and intercept was examined. The intercept was 96.53, indicating that the mean OQ-Total Score for the 110 off-track-cases was above 96 when it was determined the patient went off track. The slope before the session at which the patient went off track was 0.847 indicating that the OQ-Total Score was increasing at this rate each session. This value

was significantly different from zero ($p = 0.048$), indicating that the off track cases were deteriorating (elevating slope implying more symptoms). The slope following the first time the patient went off-track was -0.477 ($p = .014$), indicating that patients significantly improved after this point. The difference between the slopes before and after the first warning signal gives, however, limited information, as determination that a patient is not-on-track indicates an extreme score and regression toward the mean could explain decreasing scores after the signal.

The second model was conditioned on whether feedback, including the warning signal, was received or not. In the FB condition, the therapist and patient received the warning signal, where in the NFB condition they were unaware that the patient had gone off-track. Slopes were tested as an interaction with the FB condition. The slopes of the pre-signal OQ-Total Scores were not significantly different in the two conditions ($p = .327$), nor were the slopes after the signal ($p = .987$), although the effect was in the expected direction as the rate of improvement for the FB patients was greater than for the NFB patients. Therefore, it could not be conclusively demonstrated that receipt of a warning signal for cases in the FB condition resulted in better outcomes than in cases unaware of being off track.

When checking the OQ-Total Scores of 124 patients with eight or more sessions, no difference in effect of feedback between the on-track and the not-on-track cases was found when patients from both groups were included, yielding 65 in the FB condition and 59 in the NFB condition. At the third session, there was a significant difference between conditions ($t(123) = 2.01$, $p = .046$, $d = 0.36$). At the eighth session the difference between FB and NFB was only slightly greater ($t(123) = 2.21$, $p = .029$, $d = 0.40$).

To test the hypothesis that initial severity was related to the effect of feedback, we performed a multiple regression analysis with the last OQ-Total Score as the dependent variable. The independent variables were (a) the initial OQ-Total Score, (b) condition (FB

versus NFB), and (c) the interaction of the first OQ-Total Score and condition, which is the test of whether the effect of feedback depends on initial severity. The interaction effect did not produce a significant effect, although the anticipated direction and approached significance (standardized beta = 0.37, $t = 1.58$, $p = .056$, one tailed), indicating a possibility that the greater the initial distress, the larger the effect of feedback.

At the end of the study, we conducted a survey of participating therapists to determine what they perceived to be the most important aspects of the feedback system. In response to the item “How often did you open the feedback reports?” the mean was 4.47 (SD = 0.84), indicating that the responding therapists reported using the system. As well, they seemed to find it useful (mean response to “Overall Usefulness” was 4.00, SD = 0.67). The therapists found that the patient’s progress graph was the most important aspect of the OQ®-Analyst system, followed by, in order of importance, the discussion with the patient, the risk questions, comparing the patient progress with the expected progress, the signal, and the computer generated feedback message. All means were above the midpoint of the scale, but the graph and discussion were more valued than the signal and the generated message.

Conclusions: The use of OQ®-Analyst should be recommended in psychotherapeutic settings in Norway, as it in this study has shown to be an effective instrument to measure and improve outcome of psychotherapy, with significant effect already at the third session. It seems also to be useful, and perhaps especially, with severely distressed patients. In this trial, receiving the signal did not seem to be the most important therapeutic factor. Rather it appears that providing the graph to therapists and the therapist’s discussion with the patients appeared to be important.

5.0 Discussion

5.1 Main results

The OQ®-45.2 was found to be psychometrically reliable and valid for the purpose of monitoring and evaluating patient progress in therapy. It was found that feedback improved the outcome and the quality of mental services. It could not be conclusively demonstrated that receipt of a *warning signal* for cases in the FB condition resulted in better outcomes than in cases unaware of being off track. It was also found that the effect of using OQ®-Analyst occurred already in the third session, and that the graphs displaying patient progress and the discussion of progress with the patient were experienced as the most important aspects of the system.

In our three papers, we have discussed several issues concerning the use of OQ®-Analyst in Norway based on our experiences, investigations, and analyses. In this chapter, I'll emphasize some of the most important issues, and add some new considerations.

5.2 The main issues from Paper 1. The psychometric properties of OQ®-45.

5.2.1 Reliability and validity.

In the paper "Psychometric properties of the Outcome Questionnaire-45.2: The Norwegian version in an international context," we found a satisfactory reliability for the Norwegian OQ-45 Total Score as well as for the subscales, as indicated by internal consistency and test-retest correlation values consonant with other international studies. It appears that translation and administration of the OQ®-45 in Norway does not affect the reliability of the scores obtained in other studies using the OQ®-45. As well, the Norwegian version used upon ordinary psychiatric samples established concurrent validity with other validated instruments. However, the mean scores of clinical and non-clinical samples, as well as the clinical cut off and the Reliable Change Index, vary widely across national studies that have investigated properties of the OQ®-45 (Beretvas, Kearney, & Barón, 2003; Bludworth,

Tracey, & Glidden-Tracey, 2010; Chapman, 2003; Kim, Beretvas, & Sherry, 2010; Lambert et al., 1996; Lambert et al., 2011; Mueller et al., 1998; Lambert et al., 2002; de Jong et al., 2007; Chiappelli, Coco, Gullo, Bensi, & Prestano, 2008; Lo Coco et al., 2008; Wennberg et al., 2010; Qin & Hu, 2008).

The correlation of the OQ-Total Score with the SCL-90R was weaker than those found elsewhere, but the correlations with the IIP-64C were similar to the international samples, and the correlation of the SAS with SR sub-scale was higher than expected, demonstrating that the three factors were not independent. The correlations between the subscales were perhaps higher than is desirable, but similar to the values produced in other studies (cf. Chiappelli et al., 2008; de Jong et al., 2007; Lambert et al., 2002; Umphress, Lambert, Smart, Barlow, & Clouse, 1997; Qin & Hu, 2008). The high correlation is part of the reason why it is not to recommend to use the subscales for research purposes.

5.2.2 The representativeness of the samples

Our non-clinical sample had a mean age of 46.7 years, composed of adults from a rural municipality, students, members of a fitness center in a small town, health institutions employees throughout Norway and students. The sampling of the rural residents was conducted by posting OQ-45 Questionnaires in 300 random mailboxes with information emphasizing the inclusion criteria: Age above 18 years and not having a psychiatric diagnosis or being in current psychiatric treatment. Young people perhaps would have preferred an online solution instead of a letter, and could be underrepresented in this sample. On the other hand, we also collected a student sample, and there was no significant difference among our four samples. The mean age of the populations used in international psychometric studies vary for 21.6 and up to 44.3, and the sample sizes varied from 75 to 1920. It can be assumed that our sample of 338 non-clinical adults is sufficiently representative for the purpose of these studies.

The clinical sample collected for the RCT contained samples from 4 outpatient and 2 inpatient clinics, 88% of them from the outpatient clinics. There was no significant difference in the intake scores or for the outcome of the outpatient clinics. The patients from the inpatient units had, as assumed, a higher intake score. Distribution of age, gender and diagnoses are representative for patients receiving psychiatric treatment in public health institutions in Norway.

There might be several possible reasons for the variability in means for the non-clinical and clinical samples across the different countries. First, the samples used in the various countries are considerably different. The international non-clinical samples are difficult to compare, as they contain student populations, different community samples, university employees, and health employees and professionals, with mean ages ranging from 22 to 47. The sizes of the samples also varied substantially, ranging from 75 to 1920, which affects the precision of the estimates. Our sample of 280 outpatients from 3 public outpatient clinics should be large enough to be representative.

There were also considerable differences between the clinical samples in the various countries, pure student samples to samples with both inpatient and outpatient clinics in different health institutions, with mean ages ranging from 24 to 44. The mean age of our clinical outpatient sample was 34.5 years, equivalent to the mean age of the adult psychiatric patient population in Norway (Ose, Ådnanes, & Pettersen, 2014).

The relatively high level of distress found in the Norwegian clinical samples could be understood in light of the way the Norwegian health system is designed. Primary care physicians in Norway and the community health care system treat most of the patients with psychiatric disorders; patients are strictly screened before they are referred to psychiatric outpatient clinics of the type utilized in the present study.

Clearly differences in sample composition across studies and countries might well affect clinical and non-clinical means. It is therefore evident that norms and clinical cut-off scores need to be determined for each country in which the OQ®-45 is used.

5.2.3 National experienced well-being and non-clinical means of OQ-Total Score

There is a considerable variation of non-clinical means of the OQ-Total Score which are referred across countries, from 37.5 in Norway to 61 in China. This could be due to differing samples, but we also correlated the different non-clinical means of the OQ-Total Score from six different countries (USA, Germany, the Netherlands, Italy, China, and Norway) with the respective national experienced well-being. Experienced well-being was measured by the New Economics Foundation in 2012. Over 1000 respondents in 151 countries in the world were assessed using the “ladder of life” developed by the Gallup World Poll, where the bottom rung (scale value zero) represents the worst possible life and the top rung the best possible life (scale value ten) (see New Economics Foundation, 2012). The national experienced well-being scores derived in this fashion are the most extensive worldwide survey of well-being using the same instrument in existence (New Economics Foundation, 2012). The large correlation between the mean OQ-Total Score of a normal population and national experienced well-being among countries, has demonstrated that the OQ®-45 is a valid indicator of the well-being of a population.

5.2.4 The factor structure

In our study, as well as others, the correlations among the measured, as well as latent factors, were relatively large. Additionally, although the convergent validity coefficients in the present study were satisfactory, the divergent validity coefficients were almost as large or larger. Clearly, in this and other studies, the evidence for three independent factors is not convincing. However, because the OQ®-45 is intended to be used in clinical settings, the question is whether the evidence for a three factor solution is good enough to be used

clinically across countries or whether alternative structures should be adopted in the various contexts. A recent paper investigating the psychometric properties of the Swedish version of the OQ®-45 (Strid, Lundh, Andersson, & Öjehagen, 2014), concluded that the results for the different subscales were weaker, which is in accordance with our results and other studies. The less than desirable properties of the three subscales in our study mirrors findings in other countries, and is not problematic in the current research and clinical applications because in current studies, in the extant research, and in clinical practice, it is the OQ-Total Score that is used rather than the scores of the subscales.

In all, the results of the investigation of the psychometric properties suggest that the Norwegian version of the OQ®-45 demonstrated adequate, and sound psychometric properties, is reliable and valid, and adequate for research and measuring patient progress in usual care in Norway, particularly with regard to the OQ-45 Total Score.

5.3 The main issues from Paper 2. The effect of using the OQ®-Analyst in Norwegian psychiatric clinics.

The hypothesis that feedback would improve outcomes in psychotherapy was corroborated in a randomized clinical trial that was conducted in a naturalistic setting in Norway.

5.3.1 Difference between sites

It would be expected that the outcome of psychotherapy would differ in the various clinics, given the differences in severity, diagnoses, treatment models, and setting. The main effect from the site variable upon the change in OQ-45 Total Score was statistically significant, indicating that the patients in some clinics exhibited more change than others, which is understandable given that some clinics were outpatient, some inpatient, and one was a specialty outpatient clinic treating alcohol abuse. However, the *interaction* of clinic and feedback was not significant, and importantly was negligible, indicating that the effect of

feedback was not significantly different among the various clinics. Not surprising as well, there was a main effect for not-on-track cases, with not-on-track cases showing poorer outcomes than on-track cases. But the effect of *feedback* was equal for both groups (i.e., no interaction between on-track versus not-on track and feedback).

5.3.2 Does feedback influence the number of sessions?

We investigated whether the use of OQ®-Analyst had an impact upon the number of sessions. The hypothesis was that feedback would reduce the mean session number. In the present study this effect did not occur. The mean number of sessions in the FB group was 9.7, and 10.3 in the NFB group. The fact that feedback had no impact the mean session number could be interpreted in different ways. It could be that the OQ®-Analyst feedback helps the therapists to optimize the treatment length, (session number), i.e. not to prolong the therapy when recovery is achieved, and to keep the patient in therapy and reduce the drop-out rate when recovery is not achieved (Shimokawa et al., 2010), and that these two mechanisms balance each other.

5.3.3 On-track versus not-on track cases.

The effect of feedback was not significant different for on-track (not receiving any warning signals) and not-on-track (receiving one or more warning signals) patients. This is divergent from what is found in the majority of other OQ-45 studies which have found only effect for the not-on-track cases (Crits-Christoph et al., 2012; Harmon et al., 2007; de Jong et al., 2012; Lambert et al., 2001; Lambert et al., 2002; Probst et al., 2013; Simon et al., 2012; Slade et al., 2008; Whipple et al., 2003).

This result inspired us to investigate the effect of the warning signals more thorough, and also look into at which point in therapy the effect of feedback appears significant, which turned out as the focus of paper 3.

5.4 The main issues from Paper 3. How and When Feedback works: Is it the Signal?

5.4.1 The high proportion of not-on-track cases

In the present study the proportions of not-on-track cases were high in both conditions, higher than in most of the comparable studies. It seems to exist a tendency that studies in which patients had higher levels of distress (Hawkins 2004, Simon 2013, and Amble 2014), also had a higher proportion of not-on-track cases (Table 4).

Table 4. The proportions of NOT cases, Initial OQ-Total Scores and Number of sessions

Study	FB NOT%	NFB NOT%	Chi square	p	Mean Sesssion Number	Initial OQ-Total Score Full Sample	Initial OQ-Total Score On-track	Initial OQ-Total Score Not-on-track
Lambert et al., 2001	11	10	0,16	0,69		69	68	78
Lambert et al., 2002	22	25	0,92	0,34		Ca 71	68	81
Whipple et al., 2003	30	27	0,35	0,55		71	68	81
Hawkins et al., 2004	58	50	0,26	0,61	8.2	85	83	89
De Jong et al., 2012	17*	16*	0,02	0,90	16	76.7		
Crits-Christoph et al., 2012	39	38	0,02	0,88	6.5/9	69		77?
Simon et al., 2012	59**	53**	0,38	0,54	5/9	?	?	88/89
Simon et al., 2013	55	52	0,05	0,82		93/98		
Probst et al., 2013	18**	16**	0,18	0,67	6			88
Amble et al., 2014	41	44	0,12	0,73	9.9	92.3	91.7	93
Total	29	29	0,14	0,71				

*Number of Signal cases calculated based on the information of 16%.

**Number of FB and NFB cases calculated as 50% in each condition. Only NOT cases.

This is difficult to explain, as the trajectories are adjusted for the initial severity, which should ensure that approximately the same proportion of patients at each level of initial severity are expected to go off track. One interpretation could be that the more distressed patients are more complex, and less of them are following the expected trajectory curve. Another possible explanation could be that they have more sessions and therefore more “chances” to get a warning signal, and become a not-on-track case. A third explanation pertains to the fact that the trajectories for the most severe cases are based on the midpoint of a larger interval of initial scores and thus applies most accurately to patients at the midpoint of that interval; patients whose initial severity is above the midpoint will have a great probability of going off track than those below the midpoint. Thus more severely distressed samples will likely have greater likelihood of going off track. It is also interesting to notice that feedback does not reduce the number of not-on-track cases, as there is no significant difference in the number of not-on-track cases in the FB and the NFB condition in any of the studies (Table 4).

5.4.2 Effect of the warning signal

The evidence from the present RCT suggests that the warning signal itself did not have any main effect on the patient progress, as progress after the signal was not significantly different in the two conditions (FB versus NFB). This result is contrary to Propst et al. (2013), who examined the effect of the warning signal with piecewise multilevel models in the same way as performed in this study, and two other studies that examined the effect of feedback for signal cases (Crits-Christoph et al., 2013; Simon et al., 2012). The analysis of the difference of the slopes after the first warning signal shows that there is a trend toward an effect favoring the feedback condition and greater statistical power may have detected an effect. However, it is important to note that the present study had a number of off-track patients 2.5 times as great as was the case for Probst et al. (Table 4), who did detect an effect.

An important difference between the current study and Probst et al. is that the latter used Clinical Support Tools (CST), whereas the present study did not.

The result was unexpected and surprising, as it is claimed that the purpose of feedback research is to maximize outcomes for predicted treatment failures. We would have expected that receiving a warning signal alerting the therapist on a possible treatment failure, would have increased the feedback effect, but found no proof of that in the present study.

5.4.3 Initial distress level

There are differences between the present study and several of the other studies, in terms of initial distress level and the nature of the services (inpatient and outpatient, mental health versus college student samples, e.g.). The conjecture that feedback works most effectively with patients who are more distressed was supported by a trend seen in this study. Generally, when a warning signal is given, patients are relatively more distressed, which might explain why feedback is effective only in the not-on-track cases in studies with a generally lower initial level of distress (e.g., Lambert et al., 2001; Lambert et al., 2002; Crits-Christoph et al., 2012). Moreover, studies that have shown an effect overall (i.e., with the entire sample and not just the not-on-track cases) have been conducted with relatively more distressed patients in both the on-track and not-on-track conditions (viz., the current study, Hawkins et al., 2004; Simon et al., 2013). The role of initial distress is ambiguous, but the results in our study suggest that monitoring patient progress is even more important for more distressed patients, whether or not they are not-on-track cases.

5.4.4 Early effect of feedback

An important finding in the present study was that the effects of feedback occurred early, with significant differences achieved by session three. This appears to be the first study to detect and report effects for feedback so early in treatment. Interestingly, even with the early effects of feedback, the rates of off-track cases were not different in the two conditions,

meaning that feedback did not seem to reduce the number of patients going off track (See Table 4).

5.4.5 Effect of feedback for not-on-track versus on-track cases.

Prior research is not clear whether the effects of feedback are restricted to not-on-track cases or not, as studies have produced divergent conclusions. The fundamental issue here is whether the warning signal itself makes a difference, which was not found in the present study in contrast to several other studies. It could well be that some therapists are not considering the feedback until the first warning signal is given, which might be typical in high case load contexts where therapists see a large number of patients and find it difficult to focus on the difficulties of a particular patient. Another aspect might be the fact that therapists are not very good at estimating the course of therapy, especially when the patient deteriorates (Lambert, 2015; Walfish, McAlister, O'Donnell & Lambert, 2012). A warning that a given patient is not progressing may be a “wake up call” to make the therapists more attentive to this special case than they had previously been. Such a perspective is consistent with reports that many therapists do not use the feedback information, and that feedback seems to work only, or to a greater extent, for those therapists who believe feedback is useful (de Jong et al, 2012). This result is also consistent with the observation that therapist’s expertise does not seem to increase as a function of receiving feedback (Crits-Christoph et al., 2010; Tracey Wampold, Lichtenberg, & Goodyear, 2014).

5.4.6. Therapist survey

In the present study, both the patients and therapists were involved with the feedback from the first session and the therapists were encouraged to discuss the feedback reports with the patients. Therapists indicated that these discussions, as well as the patient’s progress graphs, were the most useful aspects of the feedback system, which is consistent with the relatively early effects of feedback detected in the present study. Interestingly, the PCOMS,

which explicitly involves discussing progress with the patient (Duncan, 2012), has demonstrated effects across all patients (e.g., Anker et al., 2009; Reese, Norsworthy, & Rowlands, 2009; Reese, Toland, Slone, & Norsworthy, 2010), although it is difficult to compare the two systems because of fundamental differences in the measures, procedures for using the measures, algorithms for determining signals, and in the feedback messages (see Lambert et al., 2011). In any event, it appears that in this implementation of the OQ®-Analyst, therapists and patients are using the feedback early and regularly in therapy, and there is no significant difference in effect of feedback in the FB versus the NFB condition.

5.4.7 The OQ®-Analyst system

The OQ®-Analyst has proved to be a useful and effective tool to improve outcome in psychotherapy in many countries. Also in our study, it has been experienced as useful and helpful to relate to the progress curve, the risk questions, the clinical cut-off, and the reliable change index.

The expected recovery curves are calculated from a large sample of 11000 patient courses from different psychiatric settings in US. However, much of this sample was collected from student health institutions, where the patients were less severely distressed than those included in our study. Both therapists and patients have quite often expressed that the feedback messages in the report do not fit in the patient`s actual situation. This could be due to translation errors. The messages are translated verbatim to the extent possible. Perhaps one would have emphasized a more culture sensitive translation. Another possible interpretation is that the messages are not well enough adjusted to more severely distressed patient courses. However, in our study we have asked the therapists about aspects of the system and it may well be that the messages are not found to be useful more generally.

The expected recovery curves are calculated from the initial score, and are not adjusted during therapy. A stochastic process could be employed where trajectories are

recalculated at every session, given the progress to that point. This would be one possible way to create/design a more sophisticated and flexible system.

5.5 Strengths and Limitations

Strengths of this study are the multisite design with ordinary psychiatric patients, the sample size, and the low attrition rate. We did not establish control routines to secure the therapists' adherence with the protocol. This can be regarded as a strength as this is the naturalistic setting using the system as it would be in usual care, but also as a limitation as we don't know how compliant the therapists were when it comes to opening and using the reports. De Jong et al. (2012) found that only patients with therapists who opened and used the reports benefited from feedback. On the other hand, the site where the project leader and the main project coordinator were employed and daily involved with the therapists, did not produce greater feedback effects relative to sites with more limited resources and less leader involvement. The variety of clinics increases the generalizability of the results. We know quite well how the feedback is used at one of the sites where all the therapists attended a monthly OQ-meeting, but less how it was used on the other sites where the attendance to these meetings was low or almost zero. Of course, the lack of differences in the effects of feedback across clinics supports generalizability.

There are also limitations to the present study. First, as is the case in all feedback trials, blinding is not possible. In the present study that utilized a crossed design (i.e., therapists had patients in both conditions), the therapist was well aware of whether a particular patient was in the feedback condition or not. As well, patients were informed of their assignment and were involved in reviewing reports, when their therapists followed the feedback protocol. Consequently, the effects of feedback may have been due to a Hawthorne effect (Jones, 1992) of being in the experimental condition.

Second, the heterogeneity of the clinics involved, and the relatively high distress of the patients, could limit the generalizability of the results to Norwegian conditions. It is difficult to situate the results of the present study in the array of previous results given the dimensions on which studies differ. Despite the desire to understand how feedback changes therapist behavior with patients, we were limited to the questionnaire given to the therapists who reported their opinion about what was useful, a limitation further exacerbated by the relatively low response rate and a probable bias toward therapists who had a positive attitude toward using feedback in their practice. Given these limitations and inconsistent results of previous studies, definitive conclusions should be avoided.

Part of the take home message may be that these therapists in these settings with these patients may respond as well or better to general feedback about client progress as they do to explicit signals. Setting up feedback systems is partly about getting therapists to attend to what they are doing in a different way. We also are interested in knowing the degree to which the therapists shared and discussed information about patient progress and whether these discussions were responsible for the feedback effects.

6.0 Conclusions

1. The Norwegian version of OQ®-45 shows sound and adequate psychometric properties.
2. The correlation between the non-clinical means of the OQ®-45 for individual countries and the national experienced well-being was remarkably large, which could suggest that the OQ®-45 is a valid indicator of psychological distress and well-being of a population.
3. The effect size for the difference in outcome between FB and NFB of 0.32 is moderate, in the same range as most patient feedback outcome studies.
4. The feedback effect is similar across sites and selection of patients.

5. In our study, contrary to many other international studies, the use of OQ®-Analyst did not only show effect only for the not-on-track patients, but gave better outcome both for the on-track and the not-on-track patients.
6. For the entire sample, the effect from feedback appeared early (session three).
7. Receiving a warning signal did not change of patients' progress after the signal was given when comparing the slopes in the FB and NFB conditions.
8. Feedback seemed to be more effective with more severely distressed patients, although this result did not reach significance.
9. Therapists experience the use of OQ-Analyst to be useful. They indicated that the graphs of patient progress, and the accompanying discussion with the patient, were the most important aspects of feedback.
10. The OQ®-Analyst has proved to be a useful and effective tool to improve outcome and monitor change in psychotherapy, and could be recommended to be implemented in psychiatric clinical settings in Norway.

7.0 Implications

The effect size of feedback in the present study was moderate, but is considered good enough to recommend use of OQ®-Analyst in Norwegian psychiatric settings. The moderate benefits of feedback have to be considered against the costs of implementation, including financial costs of the system, the costs of training, including opportunity costs, the time burden, the effects on the therapists within the system, and the need for a "local champion" to be responsible for the daily running of the system (Boswell, Kraus, Miller, & Lambert, 2013).

In November 2013, after this RCT-study had been terminated, we implemented the OQ®-Analyst as a routine tool at one of the participating outpatient clinics. The litmus test will be how the outcome will be for the patients using it from that point forward. Will the

effect of feedback be at the same level as for those patients who had used the OQ®- Analyst in the RCT, or will it be smaller or larger?

Future research needs to reflect curiosity about what happens in therapy as a result of feedback. In treatment trials, it is not uncommon for sessions to be video recorded and analyzed. If the same were done in feedback trials, a record of what therapists are doing with the feedback in general and specifically when a warning signal occurs, would provide insight into the process as well as the benefit of feedback. It would also be important to know how the patients experience getting and using the feedback in collaboration with their therapists, and which components they find most useful and helpful in their work to achieve a positive outcome in psychotherapy. Based on our findings, it would also be interesting to conduct a trial comparing only receiving the trajectory curve and the progress graph, with receiving the entire OQ®-Analyst feedback report with the warning signals and corresponding messages and Clinical Support Tools. In our study, the therapists appreciated the discussion with the patient, An interesting design could be to compare a condition in which OQ®-Analyst feedback is integrated into the treatment process (as in the present study) with a condition in which the feedback is given to therapist and patient but no systematic attempt is made for this information to be part of the treatment.

References

- Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment*, 55(3-4), 521-536.
- Anker, M. G., B. L. Duncan, et al. (2009). "Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting." *Journal of consulting and clinical psychology* 77(4): 693-704.
- Barkham, M., Gilbert, N., Connell, J., Marshall, C., & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *The British Journal of Psychiatry*, 186(3), 239-246.
- Barkham, M., Hardy, G. E., & Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. Chichester, UK: Wiley Blackwell.
- Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., ... & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, 69(2), 184-196.
- Bauer, D. J. & Curran, P. J. (nd) Introduction to multilevel modeling: *SPSS Demonstration Notes*. Curran-Bauer Analytics, Chapel Hill.
- Beretvas, S., Kearney, L., & Barón, A. (2003). A shortened form of the Outcome Questionnaire: A validation of scores across ethnic groups. *Published on line by the Counseling and Mind Health Centre of Texas University*.
- Berking, M., Orth, U., & Lutz, W. (2006). How effective is systematic feedback of treatment progress to the therapists? An empirical study in a cognitive-behavioral-oriented

- inpatient setting. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 35(1), 21-29.
- Bhugra, D., & Bhui, K. (2007). *Textbook of cultural psychiatry*: Cambridge University Press.
- of the Outcome Questionnaire–45. *Psychological assessment*, 22(2), 350-355.
- Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy research*, 25(1), 6-19.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230-258.
- Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models* (Vol. 154). Sage.
- Byrne, B.M. (2009). *Structural equation modeling with AMOS. Basic concepts, applications and programming*. CRC Press.
- Chapman, J. E. (2003). *Reliability and validity of the progress questionnaire: an adaptation of the outcome questionnaire*. Drexel University.
- Chiappelli, M., Lo Coco, G., Gullo, S., Bensi, L., & Prestano, C. (2008). The Outcome Questionnaire 45.2. Italian validation of an instrument for the assessment of psychological treatments. *Epidemiologia e Psichiatria Sociale*, 17(02), 152-161.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*: Routledge
- Coombs, T., Burgess, P., & Pirkis, J. (2006). Routine Outcome Measurement in Australian Child and Adolescent Mental Health Services: The Lessons of Implementation. *In the 17th World Congress of the International Association for Child and Adolescent Psychiatry and Allied Professions Melbourne*.
- Crits-Christoph, P., Ring-Kurtz, S., Hamilton, J. L., Lambert, M. J., Gallop, R.,

- McClure, B., ... & Rotrosen, J. (2012). A preliminary study of the effects of individual patient-level feedback in outpatient substance abuse treatment programs. *Journal of substance abuse treatment, 42*(3), 301-309.
- De Jong, K., Nugter, M. A., Polak, M. G., Wagenborg, J. E. A., Spinhoven, P., & Heiser, W. J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology & Psychotherapy, 14*(4), 288-301.
- De Jong, K., Timman, R., Hakkaart-Van Roijen, L., Vermeulen, P., Kooiman, K., Passchier, J., & van Busschbach, J. (2014). The effect of outcome monitoring Feedback to clinicians and patients in short and long-term psychotherapy: A randomized controlled trial. *Psychotherapy Research*
- De Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research, 22*(4), 464-474.
- Derogatis, L. (1994). SCL-90-R Symptom Checklist-90-R administration, scoring and procedures manual. Minneapolis, MN: National Computer Systems: Inc.
- Derogatis, L. R. (1977). SCL-90-R: Administration, scoring and procedures Manual I for the revised version of other instruments of the Psychopathology Rating Scale series. *Baltimore: John Hopkins University.*
- Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: an introductory report. *Psychological medicine, 3*, 595-605.
- Duncan, B. L. (2012). The Partners for Change Outcome Management System (PCOMS): The Heart and Soul of Change Project. *Canadian Psychology/Psychologie canadienne, 53*(2), 93-104.

- Durlak, J. A. (2009). "How to select, calculate, and interpret effect sizes." *Journal of Pediatric Psychology* 34(9): 917-928.
- Edwards, D. W., Yarvis, R. M., Mueller, D. P., Zingale, H. C., & Wagman, W. J. (1978). Test-Taking and the Stability of Adjustment Scales Can We Assess Patient Deterioration? *Evaluation Review*, 2(2), 275-291.
- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE—OM. *The British Journal of Psychiatry*, 180(1), 51-60.
- Evans, J. M.-C., Frank Margison, Michael Barkham, Kerry Audin, Janice Connell, Graeme McGrath, C. (2000). CORE: Clinical outcomes in routine evaluation. *Journal of Mental Health*, 9(3), 247-255.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. 2. (2004). Linear mixed effects model. *Applied Longitudinal Analysis. Probability and Statistics*, 187-233.
- Flaherty, J. A., Gaviria, F. M., Pathak, D., & Mitchell, T. (1988). Developing instruments for cross-cultural psychiatric research. *Journal of Nervous and Mental Disease*, 176(5), 260-322.
- Fonagy, P., Roth, A., & Higgitt, A. (2005). The outcome of psychodynamic psychotherapy for psychological disorders. *Clinical neuroscience research*, 4(5), 367-377.
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, 61(2), 155-163.
- Hansen, N. B., & Lambert, M. J. (2003). An evaluation of the dose–response relationship in naturalistic treatment settings using survival analysis. *Mental Health Services Research*, 5(1), 1-12.

- Hansen, N. B., Lambert, M. J. , & Forman, E. M. (2002). The Psychotherapy Dose-Response Effect and Its Implications for Treatment Delivery Services. *Clinical Psychology: Science and Practice*, 9(3), 329-343.
- Harmon, S. C., Lambert, M. J., Smart, D. M., Hawkins, E., Nielsen, S. L., Slade, K., & Lutz, W. (2007). Enhancing outcome for potential treatment failures: Therapist–client feedback and clinical support tools. *Psychotherapy Research*, 17(4), 379-392.
- Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K. L., & Tuttle, K. C. (2004). The therapeutic effects of providing patient progress information to therapists and patients. *Psychotherapy Research*, 14(3), 308-327.
- Horowitz, L., Alden, L., Wiggins, J., & Pincus, A. (2000). IIP-64/IIP-32 professional manual. *San Antonio, TX: Psychological Corporation*.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of interpersonal problems: psychometric properties and clinical applications. *Journal of consulting and clinical psychology*, 56(6), 885-892.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist*, 51(10), 1059-1064.
- Hu, L.T., Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424-453.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of consulting and clinical psychology*, 59(1), 12-19.
- Jones, S. R. G. (1992). Was there a Hawthorne effect? *American Journal of Sociology*, 451-468.

- Jöreskog, K., & Sörbom, D. (1993). LISREL 8: The SIMPLIS command language. *Chicago: Scientific Software.*
- Jöreskog, K. G. (2005). Structural equation modeling with ordinal variables using LISREL: Technical report, Scientific Software International, Inc., Lincolnwood, IL.
- Kim, S. H., Beretvas, S. N., & Sherry, A. R. (2010). A validation of the factor structure of OQ-45 scores using factor mixture modeling. *Measurement and Evaluation in Counseling and Development, 42*(4), 275-295.
- Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology, 61*(3), 285-314.
- Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future. *Psychotherapy, 52*(4), 381-390.
- Lambert, M.J., & Ogles, B. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (pp. 139-193): Chichester, John Wiley & Sons.
- Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology & Psychotherapy, 3*(4), 249-258.
- Lambert, M. J., Hannöver, W., Nisslmüller, K., Richard, M., & Kordy, H. (2002). Fragebogen zum Ergebnis von Psychotherapie. *Zeitschrift für klinische Psychologie und Psychotherapie, 31*(1), 40-46.
- Lambert, M.J., Kahler, M., Harmon, C., Burlingame, G. M. (2011). Administration and Scoring Manual Outcome Questionnaire OQ-45.2. (*OQ Measures L.L.C.*)

- Lambert, M., Morton, J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., & Burlingame, G. (2004). OQ-45.2 (Outcome Questionnaire). *American Professional Credentialing Services, LLC*.
- Lambert, M. J., & Shimokawa, K. (2011a). Measuring client outcome. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness (2nd ed.)*. (pp. 203-223). New York, NY US: Oxford University Press.
- Lambert, M. J., & Shimokawa, K. (2011b). Collecting client feedback. *Psychotherapy*, *48*(1), 72-79.
- Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice*. Washington, DC US: American Psychological Association.
- Lambert, M. J. (2013). *Bergin and Garfield's handbook of psychotherapy and behavior change*: John Wiley & Sons.
- Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., & Goates, M. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology & Psychotherapy*, *9*(2), 91-103.
- Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., & Nielsen, S. L. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research*, *11*(1), 49-68.
- Lo Coco, G., Chiappelli, M., Bensi, L., Gullo, S., Prestano, C., & Lambert, M. (2008). The factorial structure of the outcome questionnaire-45: a study with an Italian sample. *Clinical Psychology & Psychotherapy*, *15*(6), 418-423.
- Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., ... & Iveson, S. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of consulting and clinical psychology*, *73*(5), 904-913.

- Miller, S. D., Duncan, B. L., & Hubble, M. A. (2005). Outcome-Informed Clinical Work. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration (2nd ed.)*. (pp. 84-102). New York, NY, US: Oxford University Press.
- Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The partners for change outcome management system. *Journal of Clinical Psychology, 61*(2), 199-208.
- MacCallum, R.C., Browne, M.W., Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149.
- Mohr, D. C. (1995). "Negative outcome in psychotherapy: A critical review." *Clinical Psychology: Science and Practice 2*(1): 1-27.
- Monsen, J. T., Hagtvet, K. A., Havik, O. E., & Eilertsen, D. E. (2006). Circumplex structure and personality disorder correlates of the Interpersonal Problems Model (IIP-C): Construct validity and clinical implications. *Psychological assessment, 18*(2), 165-173.
- Mueller, R. M., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment, 70*(2), 248-262.
- New Economic Foundation (2012). The Happy Planet Index: 2012 Report: A global index of sustainable well-being. Retrieved July 15, 2013 from <http://www.happyplanetindex.org/assets/happy-planet-index-report.pdf>
- Ose, S. O., Ådnanes, M., & Pettersen, I. (2014). Polikliniske pasienter i psykisk helsevern for voksne 2013. *SINTEF Rapport*.
- Pedersen, G., & Karterud, S. (2004). Is SCL-90R helpful for the clinician in assessing DSM-IV symptom disorders? *Acta Psychiatrica Scandinavica, 110*(3), 215-224.

- Pedersen, G. A. (2002). Revised Norwegian version of the Inventory of Interpersonal Problems—Circumplex (IIP-C). *Tidsskrift for Norsk Psykologforening*, 39(1), 25-34.
- Pinsof, W. M., Zinbarg, R. E., Lebow, J. L., Knobloch-Fedders, L. M., Durbin, E., Chambers, A., . . . Friedman, G. (2009). Laying the foundation for progress research in family, couple, and individual therapy: The development and psychometric features of the initial Systemic Therapy Inventory of Change. *Psychotherapy Research*, 19(2), 143-156.
- Probst, T., Lambert, M. J., Loew, T. H., Dahlbender, R. W., Göllner, R., & Tritt, K. (2013). Feedback on patient progress and clinical support tools for therapists: Improved outcome for patients at risk of treatment failure in psychosomatic in-patient therapy under the conditions of routine practice. *Journal of psychosomatic research*, 75(3), 255-261.
- Qin, Y-F, & Hu, S-J. (2008). Usability Report of Outcome Questionnaire-45 in Part of Chinese Sample. *Chinese Journal of Clinical Psychology*, 16, 138-140.
- Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy: Theory, Research, Practice, Training*, 46(4), 418-431.
- Reese, R. J., Toland, M. D., Slone, N. C., & Norsworthy, L. A. (2010). Effect of client feedback on couple psychotherapy outcomes. *Psychotherapy: Theory, Research, Practice, Training*, 47(4), 616-630.
- Satorra, A., & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics & Data Analysis*, 10(3), 235-249.
- Shafran, R., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A., . . . Wilson, G. T. (2009). Mind the gap: Improving the dissemination of CBT. *Behaviour Research and Therapy*, 47(11), 902-909.

- Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*(3), 298-311.
- Simon, W., Lambert, M. J., Busath, G., Vazquez, A., Berkeljon, A., Hyer, K., . . . Berrett, M. (2013). Effects of providing patient progress feedback and clinical support tools to psychotherapists in an inpatient eating disorders treatment program: A randomized controlled study. *Psychotherapy Research, 23*(3), 287-300.
- Simon, W., Lambert, M. J., Harris, M. W., Busath, G., & Vazquez, A. (2012). Providing patient progress information and clinical support tools to therapists: Effects on patients at risk of treatment failure. *Psychotherapy Research, 22*(6), 638-647.
- Slade, K., Lambert, M. J., Harmon, S. C., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clinical Psychology & Psychotherapy, 15*(5), 287-303.
- Strid, C., Lundh, L. G., Andersson, C., & Öjehagen, A. (2014). Psychometric properties of the Swedish version of the Outcome Questionnaire-45 as administered by automated technique in a large sample of mental ill-health patients in Primary Health Care. *The European Journal of Psychiatry, 28*(4), 242-251.
- Sörbom, D., & Jöreskog, K. (1999). PRELIS 2: User's reference guide: Chicago: Scientific Software International.
- Tracey, T. J., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in Psychotherapy: An Elusive Goal? *American Psychologist, 69*, 218-229.
- Umphress, V. J., Lambert, M. J., Smart, D. W., Barlow, S. H., & Clouse, G. (1997). Concurrent and construct validity of the Outcome Questionnaire. *Journal of Psychoeducational Assessment, 15*(1), 40-55.

- Vassend, O., Lian, L., & Andersen, H. T. (1992). Norwegian versions of the NEO-Personality Inventory, Symptom Checklist 90 Revised, and Giessen Subjective Complaints List.I. *Tidsskrift for Norsk Psykologforening*, 29 (12), 1150-1160.
- Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An Investigation of Self-Assessment Bias Mental Health Providers 1. *Psychological Reports*, 110(2), 639-644.
- Walter, G., Cleary, M., & Rey, J. (1998). Attitudes of mental health personnel towards rating outcome. *Journal of Quality in Clinical Practice*, 18(2), 109-115.
- Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work*. Routledge.
- Weissman, M. M., & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. *Archives of general psychiatry*, 33(9), 1111-1115.
- Wennberg, P., Philips, B., & Jong, K. (2010). The Swedish version of the Outcome Questionnaire (OQ-45): Reliability and factor structure in a substance abuse sample. *Psychology and Psychotherapy: Theory, Research and Practice*, 83(3), 325-329.
- Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment and problem-solving strategies in routine practice. *Journal of Counseling Psychology*, 50(1), 58, 59-68.
- Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. B., Hadden, S., & Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS). Research and development. *The British Journal of Psychiatry*, 172(1), 11-18.

APPENDIX

The Norwegian version of the Outcome Questionnaire-45.

Instruksjer: Vurder siste uke, inkludert dagen i dag, slik at vi kan forstå hvordan du har hatt det. Les det enkelte utsagn nøye og kryss av under den kategorien som best beskriver din nåværende situasjon. Arbeid betyr i denne forbindelse jobb, skole, husarbeid, frivillig arbeid, daglig aktivitet osv.

Kjønn: _____ Alder: _____ Dato: _____ SETT KRYSS OVER ELLER RING RUNDT DITT ALTERNATIV:

- | | | | | | |
|--|--------|---------|-----------|---------|--------|
| 1. Jeg kommer godt overens med andre..... | Alltid | Ofte | Av og til | Sjelden | Aldri |
| 2. Jeg blir fort sliten..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 3. Jeg er ikke interessert i noen ting..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 4. Jeg føler meg stresset på arbeid/skolen/daglige aktiviteter/i min nåværende situasjon..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 5. Jeg gir meg selv skylden for mange ting..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 6. Jeg føler meg irritabel..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 7. Jeg føler meg ikke lykkelig i mitt parforhold/ nære forhold..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 8. Jeg har hatt tanker om å gjøre slutt på livet..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 9. Jeg føler meg svak..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 10. Jeg føler meg engstelig..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 11. Etter at jeg har drukket mye, trenger jeg å drikke neste dag også, for å kunne komme meg "på beina". (Hvis du ikke bruker alkohol, kryss "aldri")..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 12. Jeg er tilfreds med arbeidet mitt/skolen min/ daglige aktiviteter/ i min nåværende situasjon..... | Alltid | Ofte | Av og til | Sjelden | Aldri |
| 13. Jeg er et lykkelig menneske..... | Alltid | Ofte | Av og til | Sjelden | Aldri |
| 14. Mitt arbeid/skole/min nåværende situasjon er for mye for meg..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 15. Jeg føler meg verdiløs..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 16. Jeg er bekymret over problemer i familien..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 17. Jeg har et utilfredsstillende sexliv..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 18. Jeg føler meg ensom..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 19. Jeg kommer ofte i konflikt..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 20. Jeg føler meg elsket og satt pris på..... | Alltid | Ofte | Av og til | Sjelden | Aldri |
| 21. Jeg nyter min fritid..... | Alltid | Ofte | Av og til | Sjelden | Aldri |
| 22. Jeg har problemer med å konsentrere meg..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 23. Jeg synes fremtiden virker håpløs..... | Aldri | Sjelden | Av og til | Ofte | Alltid |
| 24. Jeg har det godt med meg selv..... | Alltid | Ofte | Av og til | Sjelden | Aldri |
| 25. Jeg har forstyrrende/ påtrengende tanker som jeg ikke klarer å kvitte meg med..... | Aldri | Sjelden | Av og til | Ofte | Alltid |

Instruks: Vurder siste uke, inkludert dagen i dag, slik at vi kan forstå hvordan du har hatt det. Les det enkelte utsagn nøye og kryss av under den kategorien som best beskriver din nåværende situasjon. Arbeid betyr i denne forbindelse jobb, skole, husarbeid, frivillig arbeid, daglig aktivitet osv.

SPØRSMÅLENE:

SCORINGENE DINE (SETT KRYSS OVER ELLER RING RUNDT DITT ALTERNATIV):

26. Jeg blir irritert over mennesker som kritiserer mitt forbruk av alkohol (eller narkotika/piller). (Hvis ikke aktuelt, kryss "aldri").....	Aldri	Sjelden	Av og til	Ofte	Alltid
27. Jeg har en irriterbar mave.....	Aldri	Sjelden	Av og til	Ofte	Alltid
28. Jeg arbeider ikke/ studerer/ fungerer ikke så godt som jeg pleide.....	Aldri	Sjelden	Av og til	Ofte	Alltid
29. Jeg har hjertebank.....	Aldri	Sjelden	Av og til	Ofte	Alltid
30. Jeg har vanskeligheter med å komme overens med venner og andre nære bekjente.....	Aldri	Sjelden	Av og til	Ofte	Alltid
31. Jeg er fornøyd med livet mitt.....	Alltid	Ofte	Av og til	Sjelden	Aldri
32. Jeg har vanskeligheter på arbeid/skole/ i min nåværende situasjon p.g.a. alkohol eller narkotika/ piller. (Hvis ikke aktuelt, kryss "aldri")...	Aldri	Sjelden	Av og til	Ofte	Alltid
33. Jeg har på følelsen at noe galt kommer til å skje.....	Aldri	Sjelden	Av og til	Ofte	Alltid
34. Jeg har ømme muskler.....	Aldri	Sjelden	Av og til	Ofte	Alltid
35. Jeg er redd for åpne plasser, kjøre bil, ta buss, trikk, T-bane, tog osv.....	Aldri	Sjelden	Av og til	Ofte	Alltid
36. Jeg føler meg nervøs.....	Aldri	Sjelden	Av og til	Ofte	Alltid
37. Jeg føler at mine kjærlighetsforhold er oppriktige og tilfredsstillende.....	Alltid	Ofte	Av og til	Sjelden	Aldri
38. Jeg føler at jeg ikke klarer meg godt på arbeid/skole/ i min nåværende situasjon.....	Aldri	Sjelden	Av og til	Ofte	Alltid
39. Jeg har for mange uoverensstemmelser på arbeid/skole/ i min nåværende situasjon.....	Aldri	Sjelden	Av og til	Ofte	Alltid
40. Jeg føler at noe er galt med "psyken" min.....	Aldri	Sjelden	Av og til	Ofte	Alltid
41. Jeg har vanskeligheter med å sovne eller sove godt.....	Aldri	Sjelden	Av og til	Ofte	Alltid
42. Jeg føler meg nedfor.....	Aldri	Sjelden	Av og til	Ofte	Alltid
43. Jeg er fornøyd med mitt forhold til andre.....	Alltid	Ofte	Av og til	Sjelden	Aldri
44. Jeg føler meg så sint på arbeidet/ skolen/ i min nåværende situasjon, at jeg kan komme til å gjøre noe jeg vil angre på.....	Aldri	Sjelden	Av og til	Ofte	Alltid
45. Jeg har hodepine.....	Aldri	Sjelden	Av og til	Ofte	Alltid

OUTCOME QUESTIONNAIRE OQ-45.2 / DEMO-KOPI / KUN TIL UNDERVISNING&FORSKNING / COPYRIGHT APOS LLC. apos@apoc.com

Welcome to OQ®-Analyst

The OQ® family of measures includes brief, self-report outcome assessment tools that are designed for repeated measurement of client progress through the course of therapy and following termination. This application is used to automate the administration and scoring of these instruments through the use of handheld PDAs, scanable paper forms, client kiosk entry, or manual data entry (10-Key). In addition, this application provides summary reports of treatment outcome in the form of ongoing client change using individualized and empirically derived classification categories (e.g., reliable improvement, no change, and reliable worsening).

OQ® Analyst can provide instant feedback to clinicians regarding patient outcomes by comparing a client's progress with the expected rate of improvement and using empirically based algorithms to predict treatment failures. A total of five clinical trials have all shown that when clinicians are informed that a patient is failing to respond to treatment in a positive manner and in accordance with expected recovery, patient outcomes are enhanced and the rate of treatment failure is greatly reduced. In essence OQ® Analyst brings to the practitioner and clinical services an automated application of the most advanced scientific and statistical methods in the field.



